NMR STUDIES OF BACTERIAL TYPE III SECRETION APPARATUS NEEDLE AND TIP PROTEINS AND THE NMR STRUCTURE OF THE HANTAVIRUS NUCLEOCAPSID COILED-COIL DOMAIN

By

2009
Yu Wang

B.S., University of Science & Technology of China (USTC), 2003

Submitted to the Department of Molecular Biosciences and the
Faculty of the Graduate School of the University of Kansas
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Committee members:

_____
Chairperson – Roberto N. De Guzman

_____
William D. Picking

_____
Wonpil Im

_____
Yoshiaki Azuma

_____
Krzysztof Kuczera

_____
Jennifer S. Laurence

Date defended: 04/08/2009

The Dissertation Committee for Yu Wang certifies
that this is the approved version of the following dissertation:

NMR STUDIES OF BACTERIAL TYPE III SECRETION APPARATUS NEEDLE AND TIP PROTEINS AND THE NMR STRUCTURE OF THE HANTAVIRUS NUCLEOCAPSID COILED-COIL DOMAIN

Committee members:

_____
Chairperson – Roberto N. De Guzman

_____
William D. Picking

_____
Wonpil Im

_____
Yoshiaki Azuma

_____
Krzysztof Kuczera

_____
Jennifer S. Laurence

Date approved: 04/22/2009

I dedicate this to my wife, my kids and my parents

Abstract

Many Gram-negative bacterial pathogens utilize type III secretion systems (TTSSs) for subverting the normal cellular functions of their target eukaryotic cells. The type III secretion apparatus (TTSA) functions like a syringe to inject proteins through an external needle and into a target cell's membrane and cytosol. The TTSA basal body spans the bacterial inner and outer membranes, and the external needle is topped with a tip complex that controls the secretion and delivery of translocator and effector proteins. The needle is formed by the polymerization of ~120 copies of a small acidic protein that is conserved among diverse pathogens. At the tip of the needle, a tip complex is assembled by tip proteins into a ring-like structure which serves as a platform for the assembly of the translocon by translocator proteins. We use NMR spectroscopy to understand how the needle is assembled and how the tip complex is assembled on top of the needle. We determined the solution structures of the BsaL needle monomer from *Burkholderia pseudomallei* and the PrgI needle monomer from *Salmonella typhimurium*. We characterized PrgI monomer-monomer interaction using NMR chemical shift mapping; and multiple contacts were found to be involved in *Salmonella* needle assembly. The tip complex is assembled by SipD, the tip protein in *Salmonella*, and BipD, the tip protein in *Burkholderia*. We also characterized PrgI-SipD and BipD-BsaL interactions by NMR. Despite weak binding affinities we learned that distinct binding sites of PrgI were involved in the PrgI-PrgI and PrgI-SipD interactions. Tip proteins were also reported to interact with

deoxycholate (DOC), a small molecule component of bile acids. We also characterized the SipD-DOC interactions by NMR. Based on data described in this dissertation, we conclude that electrostatic contacts are important in needle assembly and needle-packing interactions may be different among these bacteria. With respect to PrgI the binding sites involved in the PrgI-PrgI and PrgI-SipD interactions are also distinct. In addition, SipD-PrgI and SipD-DOC interactions provide valuable structural information to understand the activation mechanism of type III secretion.

The hantaviruses are emerging infectious viruses that in humans can cause a cardiopulmonary syndrome or a hemorrhagic fever with renal syndrome. The nucleocapsid (N) is the most abundant viral protein, and during viral assembly, the N protein forms trimers and packages the viral RNA genome. We determined the NMR structure of the N-terminal domain (residues 1–74, called N^{1-74}) of the Andes hantavirus N protein. N^{1-74} forms two long helices (α_1 and α_2) that intertwine into a coiled coil domain. The conserved hydrophobic residues at the helix α_1-α_2 interface stabilize the coiled coil; however, there are many conserved surface residues whose function is not known. Site-directed mutagenesis, CD spectroscopy, and immunocytochemistry reveal that a point mutation in the conserved basic surface formed by Arg22 or Lys26 lead to antibody recognition based on the subcellular localization of the N protein. Thus, Arg22 and Lys26 are likely involved in a conformational change or molecular recognition when the N protein is trafficked from the cytoplasm to the Golgi, the site of viral assembly and maturation.

Acknowledgements

First, I would like to thank Dr. Roberto N. De Guzman for introducing me to the exciting world of NMR and guiding me throughout my graduate career. He taught me not only how to do good science, but also how to be a good scientist. I want to thank him for his continuous support of my work and my family.

I would also like to acknowledge Dr. William D. Picking and Dr. Wendy L. Picking for their help and advice on my research. The TTSS project was initiated through collaboration with them. I am also grateful to Dr. Wonpil Im for help in computational methods and his collaboration on TTSS project, and to Daniel M. Boudreaux and Dr. Stephen C. St. Jeor, both from University of Nevada at Reno, for their help on the Hantavirus project. This work would not be possible without their help.

I would like to thank Lingling Zhang, a graduate student in Dr. Picking's lab, who worked closely with me on the BsaL project. This work could not have been completed without the help of my labmates Chet W. Egan, Dalian Zhong, Thenmalar Rathinavelan and the talented undergraduate student Nan Wang. I am also grateful to my lab colleagues: Fernando Estrada, Yan xia, Andrew Ouellette, Sunhwan Jo and Sudarsan Parthasarathy; and the undergraduate students in the lab: Arman Pirzad, Nicholas Romo, John Gallagher, and Mike Conner.

I would like to thank my committee members, and especially Drs. William Picking and Jennifer Laurence who also served as dissertation readers. I would like to

acknowledge the NMR instrumental support from the previous director of the KU NMR lab, Dr. David Van der Velde, and the current director of the biomolecular NMR lab, Dr. Asokan Anbanandam. I also acknowledge the support from the KU ITTC cluster lab.

Finally, I would like to thank my wife and colleague, Liang Zhang, who accompanies me through all these years. She is the one that is always there with me, and thanks a lot for her continued support and trust. And special thanks to my sweet daughter Sophie, my parents and my brother in China. All that I have accomplished so far wouldn't be possible without their enthusiastic support.

Table of Contents

List of Figures

List of Tables

List of Supplemental Figures

List of Abbreviations

TTSS	Type III Secretion System
TTSA	Type III Secretion Apparatus
CDC	Center for Disease Control and prevention
USDA	United States Department of Agriculture
PAI	Pathogenicity Island
SPI-1	*Salmonella* Pathogenicity Island-1
HSQC	Heteronuclear Single Quantum Coherence
HMQC	Heteronuclear Multiple Quantum Coherence
NOE	Nuclear Overhauser Effect
NOESY	NOE spectroscopy
CD	Circular Dichroism
IPTG	IsoproPyl-β-D-ThioGalactopyranoside
DMEM	Dulbecco's Modified Eagle Medium
CYANA	Combined assignment and Dynamics Algorithm for NMR Applications
AMBER	Assisted Model Building with Energy Refinement
RMSD	Root-Mean-Square Deviation
CHARMM	Chemistry at HARvard Macromolecular Mechnics
PRE	Paramagnetic Relaxation Enhancement
STEM	Scanning-Transmission Electron Microscopy

DOC	Deoxycholate
FRET	Förster Resonance Energy Transfer
TROSY	Transverse Relaxation Optimized Spectroscopy
HCPS	Hantavirus Cardiopulmonary Syndrome
N	Nucleocapsid protein
G1	Glycoprotein G1
SUMO	Small Ubiquitin-related Modifier
RNP	Ribonucleoprotein

List of Gram-negative pathogens with their associated protein components

Pathogen	Needle protein	Tip protein	Translocators
Salmonella	PrgI	SipD	SipB/SipC
Shigella	MxiH	IpaD	IpaB/IpaC
Burkholderia	BsaL	BipD	BipB/BipC
Yersinia	YscF	LcrV	YopB/YoPD
Pseudomonas	PscF	PcrV	PopB/PopD

Chapter 1: Introduction of the Type III Secretion Apparatus

THE TYPE III SECRETION SYSTEM

A common critical virulence factor for numerous Gram-negative bacterial pathogens in humans, animals and plants is the type III secretion system (TTSS).[1; 2; 3; 4] Among these bacteria are human pathogens such as enteropathogenic *Escherichia coli* (EPEC), *Salmonella* spp., *Yersinia* spp., *Shigella* spp., *Burkholderia* spp., and *Pseudomonas aeruginosa*. Illnesses caused by these pathogens range from opportunistic lung infections by *P. aeruginosa* to gastroenteritis caused by *Salmonella typhimurium*, *Shigella pestis* and *Yersinia enterocolitica*, to overwhelming systemic infections by *Yersinia pestis*, *Salmonella typhi,* and *Burkholderia pseudomallei*.[1; 2] More commonly, *Salmonella typhimurium* still elicit public health attention due to the recent outbreaks of human infections and food product recalls (www.cdc.gov). *Burkholderia* and *Yersinia* are considered serious bioterrorism threats and as CDC and USDA Select Agents.[5] In all cases, TTSS contributes to infection by allowing prokaryote communication with eukaryotic cells.[2] The outcome of this interspecies communication varies with pathogens and host cell types.[2] Nevertheless, these bacteria use TTSS for subverting normal cellular functions of their target eukaryotic cell and thereby most likely to cause infectious diseases.[1]

No vaccines are currently approved for general use against most of these pathogens.[6; 7; 8; 9] Together with the appearance of antibiotic resistant strains and the bioterrorism

potential of *Burkholderia* and *Yersinia*, these pathogens continue to pose major public health and safety concerns.[5; 10]

In *Salmonella*, TTSS-dependent invasion, survival, and persistence of pathogens within target cells are associated with damage to host tissues.[1; 11] TTSS-associated damage of host cells takes different forms, including direct cytotoxicity associated with the induction of apoptosis or necrosis or tissue damage associated with the disruption of tissue barriers.[11; 12; 13; 14; 15; 16] These mechanisms directly result in pathological outcomes, such as diarrhea, inflammation and more severe systemic infection.[17; 18]

The type III secretion system is composed of multiple protein components, which includes the type III secretion apparatus (structural proteins), translocators (structural proteins), chaperones, effector proteins and regulators (Fig. 1-1). These proteins work coordinately to fulfill the protein secretion function from the bacterial cell into the attached host cell.[19] TTSS genes are encoded on pathogenicity islands (PAIs) which have distinct G/C content compared with the adjacent DNA sequence.[20; 21] Within these islands, the *loci* organization is very similar among different bacterial species.[20; 21] However, different gene products share different levels of similarity, implying a divergence in evolution.[20; 21; 22]

While inside the bacterial cytosol, TTSS chaperonees bind specific pre-synthesized translocators and effector proteins and maintain the proteins in a secretion-competent state.[23; 24] The synthesis and secretion of late effector proteins are controlled by the regulator proteins in a temporal and spatial manner.[25; 26] Upon host

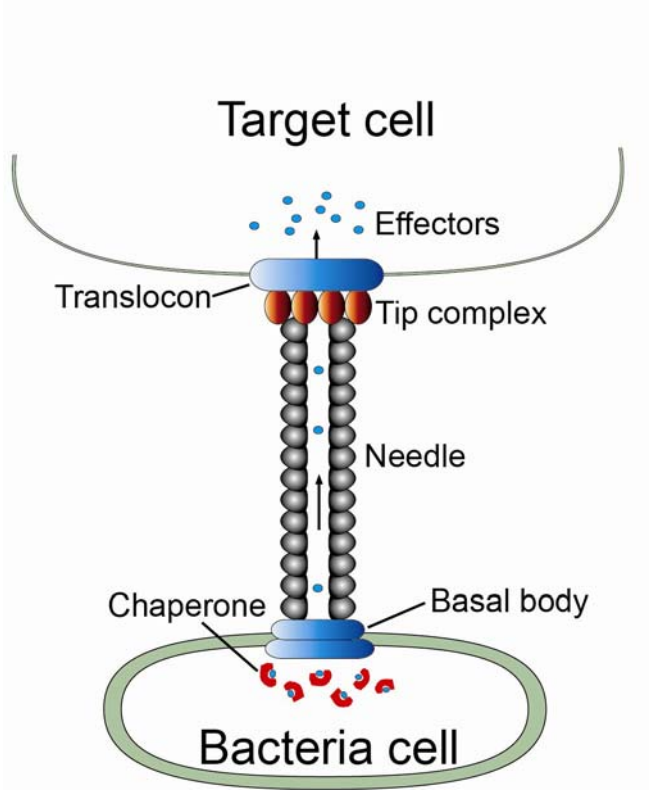

Figure 1-1. The type III secretion system (TTSS) consists of the secretion apparatus, chaperones, and effector proteins. The type III secretion apparatus (TTSA) provides a conduit for the passage of effector proteins into the cytosol of the target cell.[27]

contact, the bacterial secretion is fully activated and a subgroup of proteins is subsequently secreted in seconds.[28; 29] The protein is secreted in a very specific order, and it is believed that the secreted proteins are in a partially unfolded state.[30; 31; 32] The first secreted proteins are translocators that attach to the host cell membrane and form the translocon pore.[33; 34] The effector proteins can enter the host cell through the pore and mimic the host proteins to manipulate the host cellular function for the bacterium's benefit.[18; 35] These secreted translocators and effector proteins adopt unconserved functions and reflect host and tissue specificity.[36] In both *Salmonella* and *Shigella*, the effector proteins interact with host proteins and induce cytoskeletal rearrangments, which result in membrane ruffling that facilitates bacterial uptake, secretion of pro-inflammatory cytokines, death of macrophages or alterations in vesicle trafficking pathways inside the host cell (Fig. 1-2).[11; 18]

THE TYPE III SECRETION APPARATUS

The protein transport from a Gram-negative bacterial cell to the host cell by TTSS is powered and mediated by a specific secretion apparatus, the type III secretion apparatus (TTSA). TTSA is a macromolecular protein assembly that allows physical contact between the pathogen and its target cell.[1] TTSA is a complex protein-secretion nanomachine used to deliver bacterial effector proteins directly into target eukaryotic host cells' membrane and cytosol. These effectors then display a variety of pathogen-specific roles that alter normal host cell regulatory events.[18] Although effector proteins differ from pathogen to pathogen, TTSA, which is the

Figure 1-2. *Salmonella* uptake by human epithelial cell. Color-enhanced scanning electron micrograph showing *Salmonella typhimurium* (red) invading cultured human cells. (Fig. 1-2 is adapted from Rocky Mountain Laboratories, NIAID, NIH)

multi-protein assembly for this translocating process, is structurally and functionally conserved.[32] TTSA was first visualized by transmission electron microscopy (Fig. 1-3), and it resembles a syringe and needle (Fig. 1-1).[37] TTSA is assembled from more than 20 different proteins, most of which are associated with the bacterial inner membrane, which anchors the apparatus and provides the energy needed for the translocation of later structural units, translocator proteins and effectors in an ATP-dependent manner.[19] TTSA is composed of three major parts: the basal body, the external needle and a tip complex at the tip of the needle (Fig. 1-1).[37] The entire basal body spans the bacterial inner membrane, the peptidoglycan layer and outer membrane with the three most abundant components forming the inner and outer membrane rings.[11; 37] The extracellular portion of TTSA consists of a needle and a tip complex that controls secretion.[38]

THE TYPE III SECRETION NEEDLE

A prominent feature of TTSA is the existence of an external needle formed by the assembly of a single type of protein, the needle protein (Fig. 1-3).[39] The proper assembly of the needle is critical in bacterial pathogenesis in *Shigella* and *Salmonella*.[38; 40; 41] Analogous to a bacterial flagellum, it is believed that needle monomers pass through the channel and polymerize at the growing distal end of the needle.[32] Needle proteins are critical for bacterial virulence as needle protein null mutants are non-invasive.[38; 40; 41] Additionally, needle proteins are suggested to participate the control of the type III secretion induction.[38; 40; 41] It is also suggested

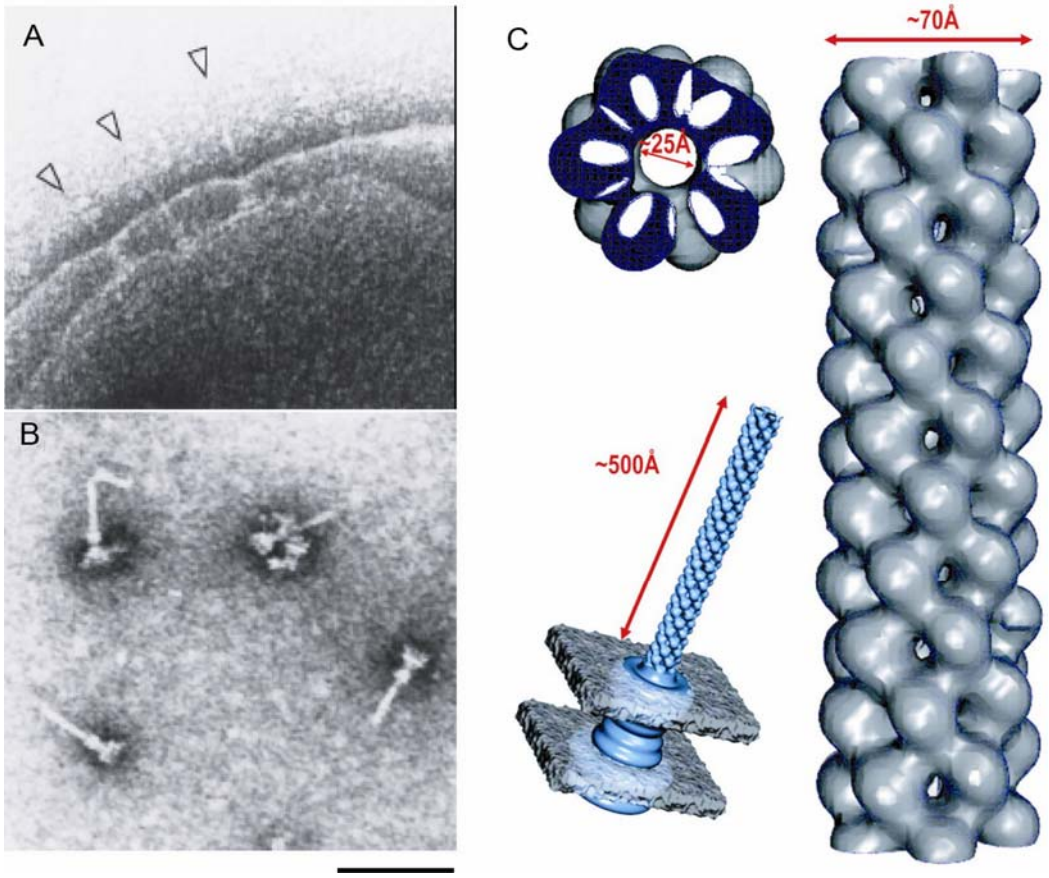

Figure 1-3. (**A**) Electron micrographs of osmotically shocked nonflagellated Δ*flhC S. typhimurium* which exhibits needle complexes on the bacterial envelope (open arrows).[37] (**B**) Needle complexes isolated obtained from an enriched fraction of the CsCl density gradient from *S. typhimurium* Δ*flhC*.[37] Scale bar, 100nm. (**C**) Various views of the three-dimensional reconstruction of *Shigella flexneri* Needle, showing the *top view*, *side view* and composite image for the structure of the needle complex in both the bacterial inner and outer membranes.[39] (Fig. 1-3A and 1-3B are from Ref. 37; Fig. 1-3C is from Ref. 38.)

that the bacterial needle mediated the transmission of the invasion signal from the outside environment into the bacterium.[42] However, detailed understanding of how the needle is assembled and how proteins travel through the needle is still not clear.

THE TYPE III SECRETION TIP COMPLEX

Bacterial translocator proteins are delivered directly into the host cell membrane to form a translocon pore through which effector proteins are then translocated into the host cytoplasm.[32] This process is mediated by a protein complex at the tip of the TTSA needle, and appropriately named tip complex.[24] The tip protein LcrV, which is also called the V-antigen of *Yersinia*, was first visualized and found to assemble into a ring-like structure at the tip of the *Y. enterocolitica* needle.[43] It is believed that the tip complex serves as a platform for the assembly of the translocon, which ultimately connects the bacterial pathogen to its eukaryotic host.[2] Tip proteins are important for bacterial type III secretion since tip proteins mutants, ***e.g.*** SipD (in *Salmonella*), IpaD (in *Shigella*) and BipD (in *Burkholderia*), showed reduced invasiveness in their respective organisms.[44; 45; 46] Although the precise sensing and triggering mechanisms of type III secretion through the tip complex are currently unknown, it is clear that direct contact between TTSA and some component of the host cell membrane leads to induction of type III secretion *in vivo*.[47] Environment-mediated secretion by the tip compelx is further transduced from the tip to the basal body through subtle conformation changes of needle packing.[42] In summary, the type III secretion apparatus (TTSA) not only controls protein transport, but also mediates

the transmission of secretion signals into the bacterium. Because the type III secretion needle apparatus is surface-exposed and critical for bacterial virulence, disrupting this protein assembly is an attractive target for the development of novel anti-bacterial agents that will prevent infection.

CONCLUSION

The increasing number of atomic structures of TTSA components have provided abundant information for understanding this complex nanomachinery.[2] Individual structures of protein components alone, however, do not answer several outstanding questions, namely: (i) how needle and tip complex are assembled, (ii) how the tip complex recognizes extracellular signals, (iii) how the invasion signal is transmitted to the TTSA base, and (iv) how the tip complex interacts with the nascent translocator proteins. More detailed understanding of protein-protein interactions and protein dynamics involved in TTSA assembly and function require a combinantion of molecular, structural and biophysical techniques. All the valuable knowledge obtained through such studies could be exploited in developing novel drugs targeted against bacterial virulence.

This dissertation mainly focuses on the characterization of the protein structures and protein-protein interactions involved in the assembly of the *Salmonella* needle and tip complex. New knowledge gained in this bacterium can be further extended to other bacterial type III secretion systems. Research in this dissertation has contributed toward understanding needle assembly by determining the atomic

structures of two needle monomers, BsaL and PrgI. I helped solving the first atomic structure of needle protein, the solution structure of monomeric BsaL, which is described in chapter 2 (adapted from a published paper by Lingling Zhang *et al*. 2006, *J. Mol. Biol.*).[48] In chapter 3, the solution structure of PrgI, the needle protein of *Salmonella* (adapted from a published paper by Yu Wang *et al*. 2007, *J. Mol. Biol.*) is described.[27] In chapter 4, the protein-protein interaction involved in the *Salmonella* needle assembly is described. In chapter 5, needle-tip and tip-deoxycholate interactions are described. Additional contents in the dissertation from chapter 1 to chapter 5 were adapted from a review paper by Yu Wang *et al*. (*Reproduced by permission of The Royal Society of Chemistry*).[2]

REFERENCES

1. Hueck, C. J. (1998). Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev* **62**, 379-433.
2. Wang, Y., Zhang, L., Picking, W. L., Picking, W. D. & Guzman, R. N. D. (2008). Structural dissection of the extracellular moieties of the type III secretion apparatus. *Mol. BioSyst.* **4**, 1176-80.
3. Van Gijsegem, F., Genin, S. & Boucher, C. (1993). Conservation of secretion pathways for pathogenicity determinants of plant and animal bacteria. *Trends Microbiol* **1**, 175-80.
4. Salmond, G. P. & Reeves, P. J. (1993). Membrane traffic wardens and protein secretion in gram-negative bacteria. *Trends Biochem Sci* **18**, 7-12.
5. Rotz, L. D., Khan, A. S., Lillibridge, S. R., Ostroff, S. M. & Hughes, J. M. (2002). Public health assessment of potential biological terrorism agents. *Emerg Infect Dis* **8**, 225-30.
6. Titball, R. W. & Williamson, E. D. (2004). *Yersinia pestis* (plague) vaccines. *Expert Opin Biol Ther* **4**, 965-73.
7. Tacket, C. O., Sztein, M. B., Losonsky, G. A., Wasserman, S. S., Nataro, J. P., Edelman, R., Pickard, D., Dougan, G., Chatfield, S. N. & Levine, M. M. (1997). Safety of live oral *Salmonella typhi* vaccine strains with deletions in htrA and aroC aroD and immune response in humans. *Infect Immun* **65**, 452-6.

8. Leary, S. E., Williamson, E. D., Griffin, K. F., Russell, P., Eley, S. M. & Titball, R. W. (1995). Active immunization with recombinant V antigen from *Yersinia pestis* protects mice against plague. *Infect Immun* **63**, 2854-8.
9. Goodin, J. L., Raab, R. W., McKown, R. L., Coffman, G. L., Powell, B. S., Enama, J. T., Ligon, J. A. & Andrews, G. P. (2005). *Yersinia pestis* outer membrane type III secretion protein YscC: expression, purification, characterization, and induction of specific antiserum. *Protein Expr Purif* **40**, 152-63.
10. Weinberger, M. & Keller, N. (2005). Recent trends in the epidemiology of non-typhoid *Salmonella* and antimicrobial resistance: the Israeli experience and worldwide review. *Curr Opin Infect Dis* **18**, 513-21.
11. Coburn, B., Sekirov, I. & Finlay, B. B. (2007). Type III secretion systems and disease. *Clin Microbiol Rev* **20**, 535-49.
12. Sato, H., Frank, D. W., Hillard, C. J., Feix, J. B., Pankhaniya, R. R., Moriyama, K., Finck-Barbancon, V., Buchaklian, A., Lei, M., Long, R. M., Wiener-Kronish, J. & Sawa, T. (2003). The mechanism of action of the *Pseudomonas aeruginosa*-encoded type III cytotoxin, ExoU. *EMBO J* **22**, 2959-69.
13. Ono, T., Park, K. S., Ueta, M., Iida, T. & Honda, T. (2006). Identification of proteins secreted via *Vibrio parahaemolyticus* type III secretion system 1. *Infect Immun* **74**, 1032-42.
14. Hauser, A. R. & Engel, J. N. (1999). *Pseudomonas aeruginosa* induces type-III-secretion-mediated apoptosis of macrophages and epithelial cells. *Infect Immun* **67**, 5530-7.
15. Finck-Barbancon, V., Goranson, J., Zhu, L., Sawa, T., Wiener-Kronish, J. P., Fleiszig, S. M., Wu, C., Mende-Mueller, L. & Frank, D. W. (1997). ExoU expression by *Pseudomonas aeruginosa* correlates with acute cytotoxicity and epithelial injury. *Mol Microbiol* **25**, 547-57.
16. Finck-Barbancon, V. & Frank, D. W. (2001). Multiple domains are required for the toxic activity of *Pseudomonas aeruginosa* ExoU. *J Bacteriol* **183**, 4330-44.
17. Zhang, S., Santos, R. L., Tsolis, R. M., Stender, S., Hardt, W. D., Baumler, A. J. & Adams, L. G. (2002). The *Salmonella enterica* serotype typhimurium effector proteins SipA, SopA, SopB, SopD, and SopE2 act in concert to induce diarrhea in calves. *Infect Immun* **70**, 3843-55.
18. Galan, J. E. (2001). *Salmonella* interactions with host cells: type III secretion at work. *Annu Rev Cell Dev Biol* **17**, 53-86.
19. Schroeder, G. N. & Hilbi, H. (2008). Molecular pathogenesis of *Shigella* spp.: controlling host cell signaling, invasion, and death by type III secretion. *Clin Microbiol Rev* **21**, 134-56.
20. Li, J., Ochman, H., Groisman, E. A., Boyd, E. F., Solomon, F., Nelson, K. & Selander, R. K. (1995). Relationship between evolutionary rate and cellular location among the Inv/Spa invasion proteins of *Salmonella enterica*. *Proc Natl Acad Sci U S A* **92**, 7252-6.

21. Hacker, J. & Kaper, J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* **54**, 641-79.
22. Perna, N. T., Mayhew, G. F., Posfai, G., Elliott, S., Donnenberg, M. S., Kaper, J. B. & Blattner, F. R. (1998). Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* **66**, 3810-7.
23. Stebbins, C. E. & Galan, J. E. (2003). Priming virulence factors for delivery into the host. *Nat Rev Mol Cell Biol* **4**, 738-43.
24. Olive, A. J., Kenjale, R., Espina, M., Moore, D. S., Picking, W. L. & Picking, W. D. (2007). Bile salts stimulate recruitment of IpaB to the *Shigella flexneri* surface, where it colocalizes with IpaD at the tip of the type III secretion needle. *Infect Immun* **75**, 2626-9.
25. Mavris, M., Page, A. L., Tournebize, R., Demers, B., Sansonetti, P. & Parsot, C. (2002). Regulation of transcription by the activity of the *Shigella flexneri* type III secretion apparatus. *Mol Microbiol* **43**, 1543-53.
26. Darwin, K. H. & Miller, V. L. (2001). Type III secretion chaperone-dependent regulation: activation of virulence genes by SicA and InvF in *Salmonella typhimurium*. *EMBO J* **20**, 1850-62.
27. Wang, Y., Ouellette, A. N., Egan, C. W., Rathinavelan, T., Im, W. & De Guzman, R. N. (2007). Differences in the electrostatic surfaces of the type III secretion needle proteins PrgI, BsaL, and MxiH. *J Mol Biol* **371**, 1304-14.
28. Bahrani, F. K., Sansonetti, P. J. & Parsot, C. (1997). Secretion of Ipa proteins by *Shigella flexneri*: inducer molecules and kinetics of activation. *Infect Immun* **65**, 4005-10.
29. Rosqvist, R., Magnusson, K. E. & Wolf-Watz, H. (1994). Target cell contact triggers expression and polarized transfer of *Yersinia* YopE cytotoxin into mammalian cells. *EMBO J* **13**, 964-72.
30. Wulf, J., Pascuzzi, P. E., Fahmy, A., Martin, G. B. & Nicholson, L. K. (2004). The solution structure of type III effector protein AvrPto reveals conformational and dynamic features important for plant pathogenesis. *Structure* **12**, 1257-68.
31. Dawson, J. E. & Nicholson, L. K. (2008). Folding kinetics and thermodynamics of *Pseudomonas syringae* effector protein AvrPto provide insight into translocation via the type III secretion system. *Protein Sci* **17**, 1109-19.
32. Blocker, A., Komoriya, K. & Aizawa, S. (2003). Type III secretion systems and bacterial flagella: insights into their function from structural similarities. *Proc Natl Acad Sci U S A* **100**, 3027-30.
33. Menard, R., Sansonetti, P. J. & Parsot, C. (1993). Nonpolar mutagenesis of the ipa genes defines IpaB, IpaC, and IpaD as effectors of *Shigella flexneri* entry into epithelial cells. *J Bacteriol* **175**, 5899-906.
34. Menard, R., Prevost, M. C., Gounon, P., Sansonetti, P. & Dehio, C. (1996). The secreted Ipa complex of *Shigella flexneri* promotes entry into mammalian cells. *Proc Natl Acad Sci U S A* **93**, 1254-8.

35. Cheng, L. W. & Schneewind, O. (2000). Type III machines of Gram-negative bacteria: delivering the goods. *Trends Microbiol* **8**, 214-20.
36. Buttner, D. & Bonas, U. (2002). Port of entry--the type III secretion translocon. *Trends Microbiol* **10**, 186-92.
37. Kubori, T., Matsushima, Y., Nakamura, D., Uralil, J., Lara-Tejero, M., Sukhan, A., Galan, J. E. & Aizawa, S. I. (1998). Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* **280**, 602-5.
38. Kenjale, R., Wilson, J., Zenk, S. F., Saurya, S., Picking, W. L., Picking, W. D. & Blocker, A. (2005). The needle component of the type III secreton of *Shigella* regulates the activity of the secretion apparatus. *J Biol Chem* **280**, 42929-37.
39. Cordes, F. S., Komoriya, K., Larquet, E., Yang, S., Egelman, E. H., Blocker, A. & Lea, S. M. (2003). Helical structure of the needle of the type III secretion system of *Shigella flexneri*. *J Biol Chem* **278**, 17103-7.
40. Kubori, T., Sukhan, A., Aizawa, S. I. & Galan, J. E. (2000). Molecular characterization and assembly of the needle complex of the *Salmonella typhimurium* type III protein secretion system. *Proc Natl Acad Sci U S A* **97**, 10225-30.
41. Blocker, A., Jouihri, N., Larquet, E., Gounon, P., Ebel, F., Parsot, C., Sansonetti, P. & Allaoui, A. (2001). Structure and composition of the *Shigella flexneri* "needle complex", a part of its type III secreton. *Mol Microbiol* **39**, 652-63.
42. Deane, J. E., Roversi, P., Cordes, F. S., Johnson, S., Kenjale, R., Daniell, S., Booy, F., Picking, W. D., Picking, W. L., Blocker, A. J. & Lea, S. M. (2006). Molecular model of a type III secretion system needle: Implications for host-cell sensing. *Proc Natl Acad Sci U S A* **103**, 12529-33.
43. Mueller, C. A., Broz, P., Muller, S. A., Ringler, P., Erne-Brand, F., Sorg, I., Kuhn, M., Engel, A. & Cornelis, G. R. (2005). The V-antigen of *Yersinia* forms a distinct structure at the tip of injectisome needles. *Science* **310**, 674-6.
44. Kaniga, K., Trollinger, D. & Galan, J. E. (1995). Identification of two targets of the type III protein secretion system encoded by the inv and spa loci of *Salmonella typhimurium* that have homology to the *Shigella* IpaD and IpaA proteins. *J Bacteriol* **177**, 7078-85.
45. Stevens, M. P., Wood, M. W., Taylor, L. A., Monaghan, P., Hawes, P., Jones, P. W., Wallis, T. S. & Galyov, E. E. (2002). An Inv/Mxi-Spa-like type III protein secretion system in *Burkholderia pseudomallei* modulates intracellular behaviour of the pathogen. *Mol Microbiol* **46**, 649-59.
46. Picking, W. L., Nishioka, H., Hearn, P. D., Baxter, M. A., Harrington, A. T., Blocker, A. & Picking, W. D. (2005). IpaD of *Shigella flexneri* is independently required for regulation of Ipa protein secretion and efficient insertion of IpaB and IpaC into host membranes. *Infect Immun* **73**, 1432-40.

47. van der Velden, A. W., Copass, M. K. & Starnbach, M. N. (2005). *Salmonella* inhibit T cell proliferation by a direct, contact-dependent immunosuppressive effect. *Proc Natl Acad Sci U S A* **102**, 17769-74.
48. Zhang, L., Wang, Y., Picking, W. L., Picking, W. D. & De Guzman, R. N. (2006). Solution structure of monomeric BsaL, the type III secretion needle protein of *Burkholderia pseudomallei*. *J Mol Biol* **359**, 322-30.

Chapter 2: NMR Structural Determination of BsaL

INTRODUCTION

A major component of TTSA is an external needle that extends from a basal structure and forms a conduit for translocating proteins into the target cell.[1] For *S. typhimurium*, TTSA needle is formed by the assembly of multiple copies of a single protein, PrgI, which is a homolog of MxiH in *Shigella*, BsaL in *Burkholderia*, YscF in *Yersinia* and PscF in *Pseudomonas*. Numerous studies have been done in *Shigella flexneri* and the exposed needle on the *Shigella* surface consists of a tube-like structure with a 70 Å outer diameter and a 25 Å inner diameter channel that extends 500 Å from the base of TTSA (Fig. 1-3).[2; 3] A similar structure exists at the surface of *Salmonella*,[4] *Yersinia*[5]and *Pseudomonas aeruginosa*.[6] Computer reconstruction of low-resolution electron microscopy images from the needles of *Shigella flexneri* shows that the approximately 120 needle monomers polymerize in a helical fashion, with 5.6 subunits per turn and a 24 Å helical pitch.[2; 3; 4]

Mutational analysis has implicated MxiH in participating in control of type III secretion induction in *Shigella* [7] with related findings reported for YscF of *Yersinia pestis*.[8] Needle proteins are critical in bacterial virulence since *mxiH* [9] and *prgI* [10] null mutants are non-invasive and do not secrete any effectors. These reports implicate TTSA needle and any associated extracellular proteins in sensing host cell contact and the transmission of type III secretion signals.[11; 12]

Sequence alignments of the needle proteins from various bacterial species show higher sequence conservation in the middle part of the proteins compared to their amino-termini (Fig. 2-1). Most of the needle proteins adopt a PxxP motif in the middle (Fig. 2-1). Secondary structure predictions indicate a predominantly α-helical secondary content with some possible β-strand structure.[7; 13] Circular dichroism (CD) spectroscopy confirms the α-helical content of the *Shigella* needle protein MxiH,[13] as well as PrgI[13] and BsaL[14]. Because of the potential of needle monomer for self-association (R. Kenjale *et al.*, unpublished results), handling recombinant needle proteins *in vitro* must be considered in such studies. Previous studies have shown that short C-terminal deletions render at least some TTSA needle proteins unable to polymerize on the bacterial surface.[7] Such deletions yielded soluble forms of MxiH[13], PrgI[13] and BsaL (R. Kenjale *et al.*, unpublished). These needle protein monomers retain all of their native secondary structure and tend to display readily reversible thermal unfolding.[13]

B. pseudomallei is the causative agent of mellioidosis, an illness that can manifest itself in humans as an acute, sub-acute or chronic infection.[15] Infection by this saprophytic bacterium is endemic to southeast Asia and northern Australia and has the potential for global spread.[15] Acute mellioidosis is a potentially fatal septicemic infection, while the sub-acute form of the disease can lead to the formation of systemic abscesses in multiple organ systems.[16] Asymptomatic infections can progress to clinical mellioidosis as is often observed in endemic regions.[17] Acute and sub-acute mellioidosis can also give rise to *B. pseudomallei* latency and the potential

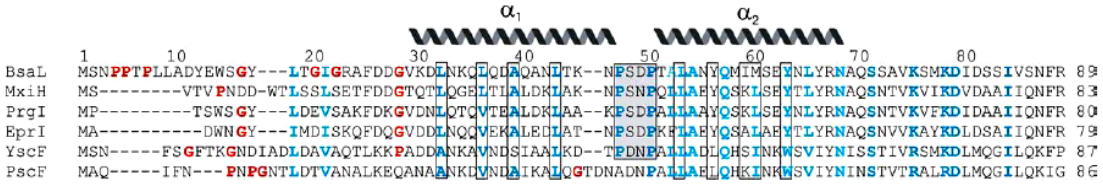

Figure 2-1. TTSS needle proteins of different bacteria show sequence conservation, especially within the core structure where helix α_1 and α_2 of BsaL are located. The conserved P-(S/D)-(D/N)-P motif that forms the inter-helical turn in BsaL is boxed. The residues at the interface of the BsaL helix α_1 and α_2, which are involved in stabilizing the structure of the core domain, are also conserved among bacterial species (also boxed). Conserved and identical residues are in blue, and the glycine and proline residues preceding helix α_1, are in red. Sequences are from Swiss-Prot, and the bacterial species and its associated needle protein are: *B. pseudomallei* (BsaL), *S. flexneri* (MxiH), *S. typhimurium* (PrgI), *E. coli O157:H7* (EprI), *Y. pestis* (YscF), and *P. aeruginosa* (PscF).[14]

for delayed, and often fatal, relapse [16] despite appropriate antibiotic therapy during the initial infection.[18] *B. pseudomallei* and its close relative *Burkholderia mallei* (causative agent of glanders) are considered important bioterrorism agents due to their ready world-wide availability and their ability to be transmitted by aerosol.[19] *B. pseudomallei* possesses at least one genetic region that encodes a TTSS, resembling that of the plant pathogens *Ralstonia solanacearum* and *Xanthomonas* spp.[19] A separate TTSS (the bsa locus) encoded within a recently identified *B. pseudomallei* pathogenicity island [19] is similar to those encoded by the *Shigella* virulence plasmid (the Mxi/Spa TTSS) and *Salmonella* pathogenicity island SPI-1 (Inv/Spa TTSS).[19] These genetic elements encode proteins such as invasins (Inv), membrane excretion of invasins (Mxi), and surface presentation of antigens (Spa), required by *Shigella* and *Salmonella* to become virulent. The importance of the Bsa TTSS for *B. pseudomallei* virulence and pathogenesis is not yet entirely clear; however, it has been predicted that cellular invasion and survival in macrophages is an important aspect of *B. pseudomallei* pathogenesis.[20] Furthermore, Stevens *et al*. showed that the Bsa TTSS is critical for *B. pseudomallei*'s ability to invade and replicate in J774.2 murine macrophages and for vacuolar escape following host cell invasion.[19] Similar to MxiH and PrgI, deletion of the last five residues of BsaL[14] prevented self-association and yielded soluble proteins that allowed biophysical studies including the NMR structural determination of $BsaL^{C\Delta5}$.

In this chapter, the NMR structure of the needle protein of *B. pseudomallei*, BsaL, lacking its five C-terminal residues ($BsaL^{C\Delta5}$) is described. $BsaL^{C\Delta5}$ is

composed of two well-defined helices withing the central portion of the protein's primary sequence, and flanked by residues that appear to adopt a partial α-helical conformation. The NMR data rule out the possibility of an internal β-strand and provide insight into the possible sites of the monomer-monomer interactions that are likely to be needed to form the *B. pseudomallei* Bsa TTSA needle. The NMR structure determination of BsaL was done as collaboration with Lingling Zhang (Dr. Picking's lab). Lingling Zhang initiated the project and prepared the NMR samples, acquired the NMR data, and assigned the backbone resonance of $BsaL^{C\Delta5}$. My primary contributaion is in the assignment of NOESY datasets, generation of restraints, building initial structural models and the NMR structure calculation and refinement of $BsaL^{C\Delta5}$.

MATERIALS AND METHODS

Protein Expression and Purification

The expression plasmid of $BsaL^{C\Delta5}$ was from Dr. Wendy L. Picking and Dr. William D. Picking. Lingling Zhang overexpressed and purified $BsaL^{C\Delta5}$. Details about protein expression and purification of $BsaL^{C\Delta5}$ can be found in Lingling Zhang's doctoral dissertation.

NMR Spectroscopy and Structure Calculation

NMR data were acquired at 25°C on a Bruker Avance 800 MHz NMR spectrometer equipped with a cryogenic triple-resonance probe. Resonance

assignments were obtained from 2D $^{1}H–^{15}N$ HSQC[21] and $^{1}H–^{13}C$-HMQC[22] and the following 3D datasets: HNCA,[23] HNCO,[24] CBCA(CO)NH,[23] ^{15}N-edited NOESY-HSQC (t_{mix} 120 ms)[25] and ^{13}C-edited HMQC-NOESY (t_{mix} 120 ms) (Table 2-S1).[26] NMR data were processed using NMRPipe[27] and analyzed using NMRView.[28] Inter-proton distance restraints were obtained from the NOESY datasets above and were classified based on peak volumes into upper bounds of 2.7, 3.5, 4.5 and 5.5 Å and lower bound of 1.8 Å. Based on the secondary chemical shifts (Fig. 2-2),[29] the backbone dihedral angles of residues in helix α_1 and α_2 were restrained to $\phi = -60(\pm 20)°$ and $\psi = -40(\pm 20)°$. Additional negative ϕ restraints were imposed on non-glycine residues based on the chemical shifts.[29] Structure calculations and refinement were performed as follows. An initial set of 200 structures were generated by torsion angle dynamics using CYANA,[30] and 100 structures with the lowest target functions were further refined by molecular dynamics and simulated annealing *in vacuo* using AMBER7 with the ff99 force field. From this, a set of 50 structures with the lowest energies were further refined in AMBER7 with the generalized Born (GB) potential. Details of AMBER calculations can be found in Dames *et al*.[31] Twenty structures with the lowest AMBER energies were analyzed by PROCHECK[32] and graphics were made using MOLMOL[33] and PyMOL[34].

RESULTS

Preparation of BsaL$^{C\Delta 5}$ for Structural Analysis

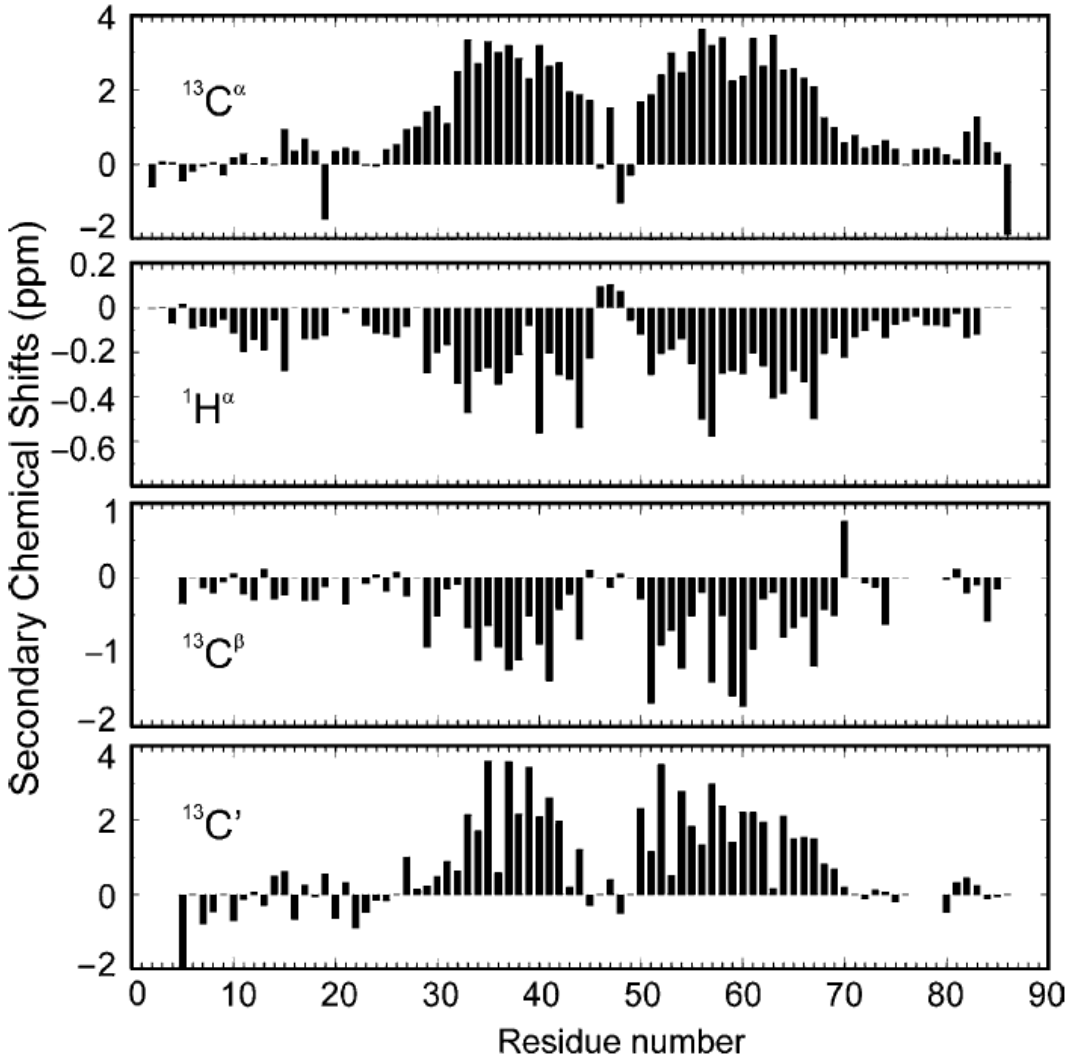

Figure 2-2. Secondary chemical shifts obtained by subtracting the random coil $^{13}C_{\alpha}$, $^{1}H_{\alpha}$, $^{13}C_{\beta}$, and $^{13}C'$ chemical shift values from those observed in the protein indicate the presence of two well-defined helices in the middle of $BsaL^{C\Delta5}$.[14]

Recombinant *B. pseudomallei* BsaL that lacks five C-terminal residues and possesses a His_6 C-terminal tag was subcloned in the NdeI/XhoI sites of pET-22b vector (Novagen) and was expressed and purified by Ni^{2+}-chelation chromatography as described.[13] CD data indicated that $BsaL^{C\Delta5}$ was rich in α-helical secondary structure (approximately 70%) and was similar, in this respect, to $MxiH^{\Delta5}$ and $PrgI^{\Delta5}$.[13] The 2D $^{1}H–^{15}N$ heteronuclear single quantum coherence (HSQC) spectrum of recombinant $BsaL^{C\Delta5}$ (Fig. 2-3) showed a narrow ^{1}H chemical shift range with well-dispersed peaks, consistent with the likelihood that this protein contained a high degree of α-helical structure. NMR structure determination was performed on $BsaL^{C\Delta5}$ because of the high degree of solubility and apparently well-defined structural content of $BsaL^{C\Delta5}$.

NMR Spectroscopy of $BsaL^{C\Delta5}$

Analysis of 3D HNCA (Fig. 2-S1),[23] CBCA(CO)NH (Fig. 2-S2),[23] HNCO (Fig. 2-S3),[24] ^{13}C-edited HMQC-NOESY,[26] and ^{15}N-edited NOESY-HSQC (Fig. 2-S4)[25] experiments (Table 2-S1) afforded near complete sequence-specific backbone and side-chain resonance assignments of $BsaL^{C\Delta5}$. Ser2 and three residues at the C terminus, Met77–Asp79, are missing from the $^{1}H–^{15}N$ HSQC spectrum due to exchange broadening and could not be assigned. The His residues in the C-terminal His_6-tag are disordered, as shown by secondary chemical shifts in the random coil

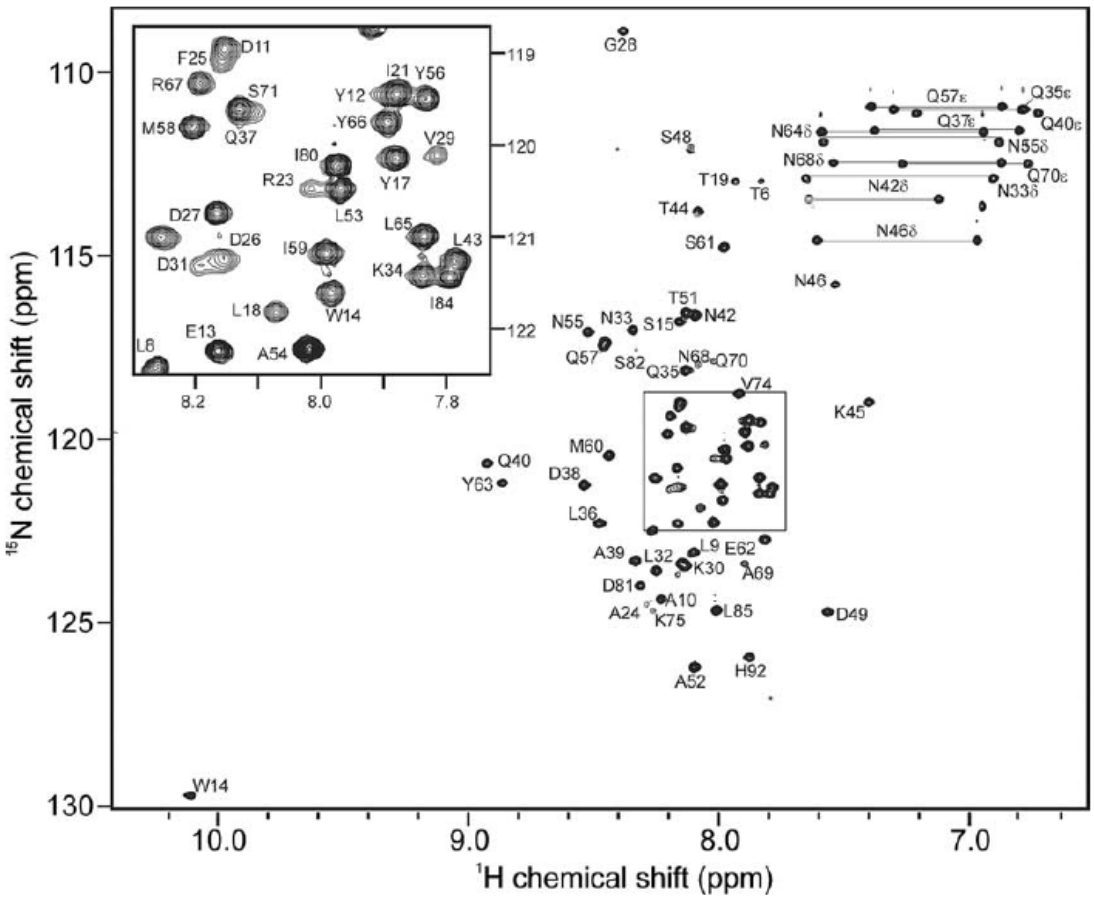

Figure 2-3. The 2D NMR ^{1}H, ^{15}N correlation spectrum (HSQC) of the recombinant ^{15}N-labeled BsaL$^{C\Delta5}$ showed well-dispersed peaks and a narrow proton chemical shift range of about 1.5 ppm, suggesting a predominantly helical structure. Assignments are indicated, and the insert shows the expansion of the central portion of the spectrum (boxed).[14]

range, and ^{15}N and ^{13}C NOESY peaks that showed only strong intraresidue nuclear Overhauser effect (NOE) peaks. Analysis of the H_α, C_α, C_β, and C′ secondary chemical shifts (Fig. 2-2) suggested the presence of two well-defined helices in $BsaL^{C\Delta5}$ (Fig. 2-2).[29] This was further confirmed by the NOE patterns,[35] such as $H_N(i)$-to-$H_N(i+2)$, (where *i*=residue number), $H_\alpha(i)$-to-$H_\beta(i+3)$, and $H_\alpha(i)$-to-$H_N(i+4)$ NOEs (Fig. 2-S5), suggesting that these helices adopt a regular α-helix conformation.

The Central Portion of $BsaL^{C\Delta5}$ Forms a Two-Helix Bundle

Structure calculation and refinement using CYANA and AMBER yielded well-converged structures (Fig. 2-4A) with low energies and low distance and dihedral angle restraint violations (Table 2-1). A total of 716 NOE-derived distance restraints and 67 φ and 33 ψ dihedral angle restraints were used in the structure calculation (Table 2-1). PROCHECK analysis of the ensemble of 20 low energy structures showed that more than 99% of the residues are within the allowed region of the Ramachandran plot (Table 2-1). The central portion of $BsaL^{C\Delta5}$, spanning from Lys30 to Asn68, forms the highly structured core domain of $BsaL^{C\Delta5}$, which has two well-defined helices joined by an ordered, four-residue turn. More than 80% of the assigned NOEs and nearly all of the long-range NOEs are found within this core domain, yielding an average of about 15 restraints per residue (Table 2-1). Helix α_1 has five helical turns and spans residues Lys30–Asn46. This is followed by a highly ordered linker composed of Pro47-Ser-Asp-Pro50, which immediately connects with the seven-turn helix α_2 spanning residues Thr51–Asn68.

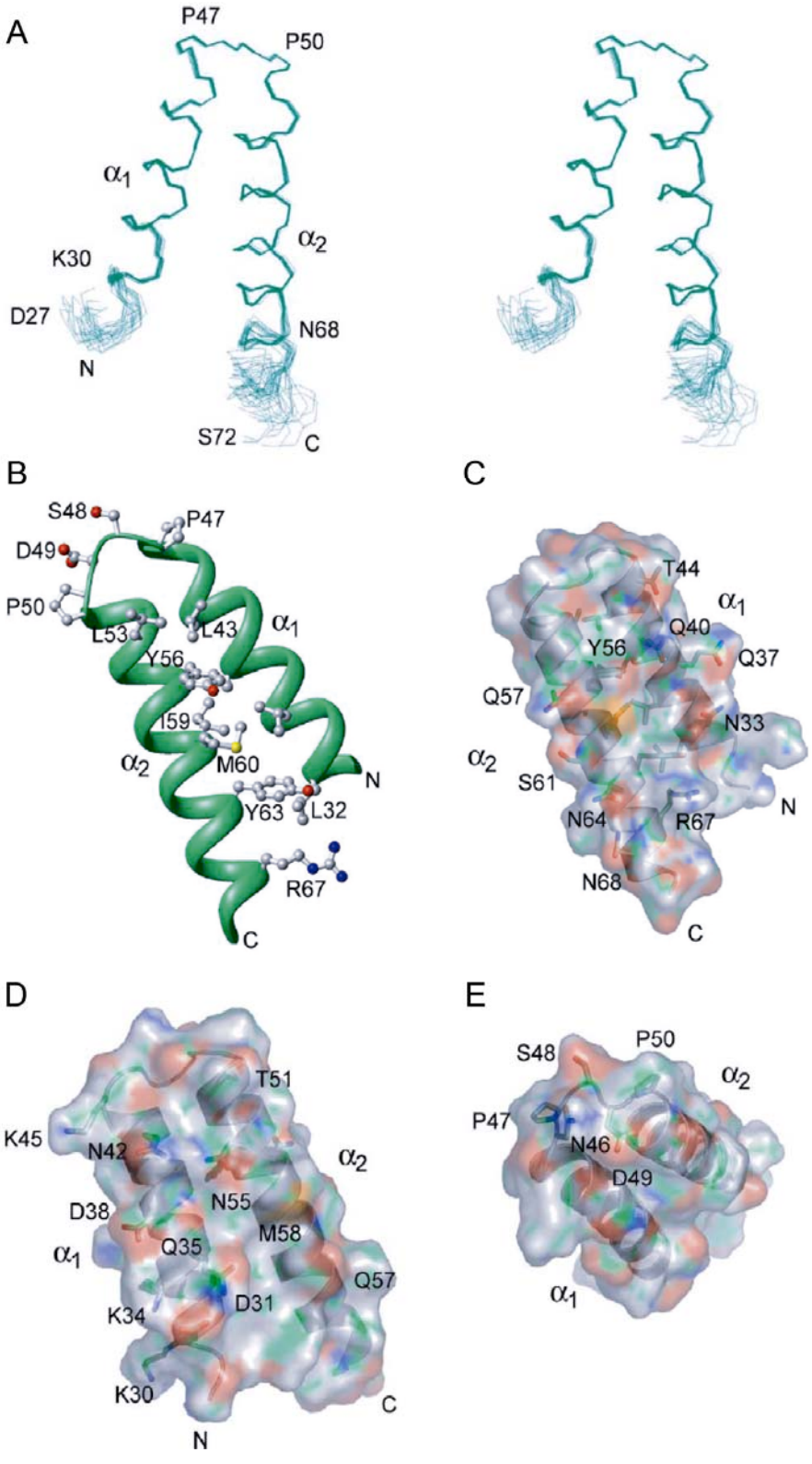
A
P47
P50
α1
α2
K30
D27
N68
N
S72
C
B
S48
P47
D49
P50
L53
L43
α1
Y56
I59
α2
M60
N
Y63
L32
R67
C
C
T44
α1
Q40
Y56
Q37
Q57
N33
α2
S61
N
N64
R67
N68
C
D
T51
K45
N42
α2
N55
D38
M58
α1
Q35
Q57
D31
K34
K30
C
N
E
S48
P50
α2
P47
N46
D49
α1

Figure 2-4. (**A**) Stereo view of the 20 lowest energy NMR structures of $BsaL^{C\Delta5}$ showing residues Asp27–Ser72. Residues Lys30–Asn68 form the well-defined core domain of $BsaL^{C\Delta5}$. The rest of the molecule does not adopt a single conformation and, therefore, will not superimpose. (**B**) Ribbon diagram of the lowest energy NMR structure of $BsaL^{C\Delta5}$, showing the residues that form the hydrophobic contacts at the interface of helix α_1 and α_2. ((**C**)–(**E**)) Surface representations showing the polar residues that line the surfaces of the core domain of $BsaL^{C\Delta5}$. The side-chain functional groups of residues on the surface are colored as follows: amides blue, carboxyls and hydroxyls red, methylenes green, and the methionine sulfur atoms orange. The amino and carboxy termini are labeled N and C.[14]

Table 2-1. Number of restraints for $BsaL^{C\Delta5}$ (residues M1-I85) and the core domain (residues K30-N68), and structure statistics for the 20 low energy structures of $BsaL^{C\Delta5}$.[14]

	K30-N68	M1-I85
A. *Number of restraints*		
Total distance restraints	582	716
Intraresidue (i,i)[a]	116	133
Sequential $(i,i+1)$	127	193
Medium range $(2 \leq \lvert i-j \rvert \leq 4)$	217	267
Long range $(\lvert i-j \rvert > 4)$	122	123
Total dihedral angle restraints	67	100
Phi	34	67
Psi	33	33
B. *Structure statistics (20 structures)*		
RMS[b] deviation from the mean structure (K30-N68)		
Backbone atoms (N,C^{α},C') (Å)	0.37	
All heavy atoms (Å)	0.88	
Violation analysis		
Maximum dihedral angle (deg.)	2.4	
Maximum distance (Å)	0.28	
Energies		
Mean AMBER energy (kcal/mol)	−3183	
Mean restraint energy (kcal/mol)	37	
Deviation from idealized geometry		
Bond lengths (Å)	0.0100 ± 0.0001	
Bond angles (deg.)	2.43 ± 0.02	
PROCHECK statistics		
Most favorable (%)	87.2	
Additionally allowed (%)	12.3	
Generously allowed (%)	0.4	
Disallowed (%)	0.1	

[a] i,j = residue number.
[b] RMS, root-mean-square.

The Two Helix bundle Is Stabilized by Hydrophobic Interaction

The BsaL$^{C\Delta5}$ core domain is stabilized by hydrophobic interactions (Fig. 2-4B), mediated by residues that are also conserved among needle proteins (Fig. 2-1). The hydrophobic side of the amphiphatic helix α_1 forms direct contact with the hydrophobic side of helix α_2. The two helices pack at an angle of 24°, and the interface is formed by the hydrophobic interactions of Leu32, Leu36, Ala39, and Leu43 on helix α_1 with Tyr63, Met60, Ile59, Tyr56, Leu53, and Ala52 on helix α_2. The aromatic rings of Tyr56 and Tyr63 are stabilized by hydrophobic contacts. The Tyr56 ring is sandwiched between the hydrophobic side-chains of Leu43 and Met60, and forms additional contacts with the methylene groups of Gln40 and Gln67. Similarly, the Tyr63 ring is wedged between Leu36 and the methylene groups of Arg67 with additional contacts to side-chains of Leu32 and Gln64 (Fig. 2-4B).

The Surface of the Core Domain Is Lined with Polar Residues

The surfaces of the core domain are dominated by polar groups. One side is lined with polar residues: Gln40, Gln37, Asn33 on helix α_1 and Gln57, Ser61, Gln64, and Asn68 on helix α_2. In addition, the hydroxyl groups of Tyr56 and Tyr63 are pointed toward the surface and contribute to the overall polarity of this side. Asn68 is conserved among needle proteins (Fig. 2-1); however, it does not participate in hydrophobic contacts at the helix α_1-α_2 interface, suggesting that it might be involved in functions other than stabilizing the core domain. The other side of BsaL$^{C\Delta5}$ is also

lined with polar residues: Lys30, Asp31, Lys34, Gln35, Asp38, Asn42, and Lys45 on helix α_1 and Thr51, Asn55, Gln57, and the hydroxyl group of Tyr66 on helix α_2 (Fig. 2-4C). The linker between helix α_1 and α_2 is also polar, comprised of Ser48 and Asp49 and close to the nearby Asn46 at the end of helix α_1 (Fig. 2-4C). The polar nature of the helix-turn-helix core domain of BsaL$^{C\Delta5}$ suggests that this domain is solvent-exposed in the needle apparatus.

Significance of Two *trans* Proline Residues of BsaL$^{C\Delta5}$

The two proline residues located at the center turn, the PxxP motif, of BsaL$^{C\Delta5}$, Pro47 and Pro50, are in *trans* conformation, as evidenced by strong NOEs from the proline H_δ protons to the H_α protons of the preceding residues, Asn46 and Asp49, respectively (Fig. 2-S6). The *trans* conformation of these two Pro residues constrains the orientation of helices α_1 and α_2, so that they must be positioned next to each other. That is, they must assume a structure that is very close to the final global fold of the core domain of the needle protein. These two Pro residues are highly conserved among TTSA needle proteins in diverse bacterial species (Fig. 2-1). The constraining action of these two Pro residues suggests a role in the folding of all these needle protein homologues and may provide an explanation for the reversible unfolding of the protein *in vitro*.[13]

Residues Flanking the Core Domain

The first seven residues of BsaL$^{C\Delta5}$ are unstructured. These residues have no apparent secondary structures from chemical shifts and show only intra-residue NOEs (Fig. 2-S5). Based on sequential NOE patterns, the three proline residues in this region, Pro4, 5, and 7, are all in *trans* conformation, with a minor *cis* conformation also observed for Pro4. The regions flanking the well-defined core domain of BsaL$^{C\Delta5}$, Ala10 to Gly28 at the N terminus and Ser71 to Ile84 at the C terminus, are partially ordered. They exhibit medium-range NOE patterns that are more characteristic of α-helical structures. For example, residues from Leu9 to Gly28 and Ser71 to Ile84 showed a consistent pattern of H_α(i)-to-H_β(i+3) NOEs, with weaker intensities compared to the corresponding H_α(i)-to-H_β(i+3) NOEs found in helix α_1 and α_2 (Fig. 2-S5). There are also sequential H_N(i)-to-H_N(i+1) NOEs found from Leu8 to Gly28, and for some of the residues between Ser71 to Ile84, again with weaker intensities compared to those found in helix α_1 and α_2. It is not possible to ascertain the presence (or absence) of weak H_N(i)-to-H_N(i+1) NOEs between Ser71 and Ile84 with the available NMR data due to spectral overlap. The NOE patterns suggest the regions flanking the core domain are not completely disordered but are more likely to be in partial-helical conformations.

No β-hairpin in BsaL$^{C\Delta5}$

In the absence of high-resolution structural data for the needle proteins of TTSA, the *Salmonella* flagellar filament FliC was used to predict the structure of the needle proteins of TTSA.[7; 13] This secondary structure prediction suggested the

presence of a β-hairpin in the needle proteins,[7; 13] corresponding to residues Tyr63-Gln70 of BsaL$^{C\Delta5}$. This is not supported by the NMR data presented here. The secondary chemical shifts, NOE patterns, and NMR structure indicate that Tyr63–Gln70 forms the C-terminal part of helix α_2 and cannot be in a β-hairpin conformation. Based on the sequence similarity, it is predicted that other needle proteins will also fail to have a β-hairpin in this region (Fig. 2-1).

Structural homology with other proteins

As expected, the two-helix bundle is a common structural feature of many proteins, and a search using DALI[36] yielded at least 52 proteins that are structurally homologous to the core domain of BsaL$^{C\Delta5}$. These proteins contain helix-turn-helix structures that can be superimposed with the core domain of BsaL$^{C\Delta5}$ with a root-mean-square deviation (RMSD) of 2.0 Å or better. Protein bearing similar fold varies from membrane proteins, enzymes, chaperones, and transcription factors found in eukaryotes, bacteria, and viruses that are involved in cell adhesion, gene regulation, signaling, and metabolism. What is noteworthy is the lack of sequence identity among these proteins. Of the 52 proteins, 42 had sequence identity below 10% with the highest identity being a designed protein (23%). Among naturally occurring proteins, the highest sequence identity was from a yeast chaperone protein (18%). There is structural homology with another TTSS protein, YscE, of *Y. pestis,*[37] which had a backbone RMSD value of 1.9 Å and 14% sequence identity with the BsaL$^{C\Delta5}$ core domain.

DISCUSSION

The bacterial type III secretion apparatus consists of a basal body that spans the inner and outer membranes of Gram-negative bacteria and an external needle formed by the polymerization of multiple copies of a single type of protein into a tube-like helical assembly (Fig. 1-1).[38] These substructures give rise to an apparatus that is capable of delivering bacterial proteins into eukaryotic cells for the subversion of normal cellular functions. Upon assembly, the needle apparatus has a 25 Å diameter channel that is used for translocating bacterial effector proteins into the host cell membrane and cytoplasm (Fig. 1-3). For high-resolution structural determination of monomeric needle proteins, it was necessary to consider potential problems that could arise from the intermolecular protein–protein interaction of recombinant needle subunits. Such issues probably reflect the lack of high-resolution structures for any TTSS needle protein in the literature. Fortuitously, deletion of the C-terminal five residues of the needle proteins MxiH from *Shigella*,[13] PrgI from *Salmonella*,[13] and BsaL[19] results in the production of monomeric forms of these proteins. Monomers produced in this way retain extensive native secondary structures, as determined by biophysical characterization,[13] thus allowing high-resolution structure determination of these needle protein monomers. The central portion of the needle proteins shows strong sequence conservation (Fig. 2-1) and all of those purified as monomers appear to have very similar secondary structures that are rich in α-helical content.[13] It was expected that structure determination for any one of the needle proteins would

provide a significant contribution to the current understanding of how needle protein structure correlates with its biophysical properties, its polymerization into type III secretion apparatus needles, and possibly its role in pathogenesis.

Recombinant BsaL$^{C\Delta5}$ showed ideal characteristics needed for NMR-based structure determination. BsaL$^{C\Delta5}$ displayed near-millimolar solubility and well-dispersed peaks in the 2D $^{1}H–^{15}N$ HSQC spectrum (Fig. 2-3), making it highly suitable for NMR structural characterization. Analysis of secondary chemical shifts clearly identified two well-defined helices (Fig. 2-2) and further analysis of the NOESY datasets yielded a good number of distance restraints (Table 2-1), which allowed for atomic-level structure determination with good structural statistics.

As a monomer, the central portion of BsaL$^{C\Delta5}$ forms a well-structured, two-helix bundle. The first seven residues are completely disordered, and the regions flanking the well-structured core domain are partially ordered. This is reflected in a higher number of distance restraints, averaging about 15 restraints per residue within the central region. In contrast, the flanking regions showed medium-range NOEs that are characteristics of α-helical structures. The NMR structure of BsaL$^{C\Delta5}$ (Fig. 2-4A) supports the CD data showing that this needle protein is predominantly α-helical and it is likely that PrgI (*Salmonella*) and MxiH (*Shigella*) have similar structures as well.[13] The data presented here do not, however, support the previous suggestion that there may be a β-hairpin within the C-terminal half of these needle proteins.[7; 13] The core domain is stabilized by hydrophobic interaction and the surface is lined with polar residues, suggesting that this part of the needle protein may be solvent exposed

in the needle apparatus. The atomic structure of $BsaL^{C\Delta5}$ could also account for associated phenotypes of some of the point mutants that have been made on the homologous needle protein of *S. flexneri*, MxiH, where a larger body of biological, biochemical and genetic data are available.[7] For example, the MxiH mutant Y57A results in a TTSS that is incapable of inducible secretion and a TTSA that forms unusually short needles.[7] This mutation most likely disrupted the hydrophobic interactions that stabilize the core domain, since the equivalent residue in BsaL, Y63, lies at the heart of the BsaL monomer helix-helix interface.

The two-helix bundle is a common structural motif in biology and numerous proteins share structural homology with the core domain of $BsaL^{C\Delta5}$, despite an absence of overt primary sequence similarity. One protein from *Yersinia*, YscE, shares 14% sequence identity and has a backbone RMSD value of 1.9 Å with the core domain of BsaL. YscE is classified in the Protein Data Bank as a type III secretion chaperone, and was reported by Phan and co-workers[37] to form dimers in solution. The homolog of YscE in *Pseudomonas*, PscE, together with another protein, PscG, forms a stable ternary complex with the needle protein, PscF, which then prevents the premature polymerization of the needle protein in the bacterial cytoplasm.[39] A functional homolog of YscE/PscE in the *Burkholderia* Bsa TTSS has not been identified.

Based on the atomic structure of $BsaL^{C\Delta5}$, additional experiments are needed to answer the question of how two BsaL monomers might interact to form an external TTSA needle. A working hypothesis is that the intermolecular contacts between the

needle molecules required for self-assembly reside within the N-terminal residues 10–25 and involves the last five residues at the C terminus of BsaL, and upon complex formation, these residues form helical structures that pack against each other. This is consistent with the model of how needle subunit proteins might polymerize based upon similarities to the flagellar structure.[3; 7; 38] The flagellar structure is assembled by intermolecular helix-helix packing between monomers.[40] The results presented herein indicate that the residues flanking the core domain are not completely disordered, but are best characterized as having partial helical character, hence protein–protein interaction involving helix–helix interaction will be thermodynamically favored.

REFERENCES

1. Kubori, T., Matsushima, Y., Nakamura, D., Uralil, J., Lara-Tejero, M., Sukhan, A., Galan, J. E. & Aizawa, S. I. (1998). Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* **280**, 602-5.
2. Cordes, F. S., Komoriya, K., Larquet, E., Yang, S., Egelman, E. H., Blocker, A. & Lea, S. M. (2003). Helical structure of the needle of the type III secretion system of *Shigella flexneri*. *J Biol Chem* **278**, 17103-7.
3. Cordes, F. S., Daniell, S., Kenjale, R., Saurya, S., Picking, W. L., Picking, W. D., Booy, F., Lea, S. M. & Blocker, A. (2005). Helical packing of needles from functionally altered *Shigella* type III secretion systems. *J Mol Biol* **354**, 206-11.
4. Marlovits, T. C., Kubori, T., Sukhan, A., Thomas, D. R., Galan, J. E. & Unger, V. M. (2004). Structural insights into the assembly of the type III secretion needle complex. *Science* **306**, 1040-2.
5. Journet, L., Agrain, C., Broz, P. & Cornelis, G. R. (2003). The needle length of bacterial injectisomes is determined by a molecular ruler. *Science* **302**, 1757-60.
6. Pastor, A., Chabert, J., Louwagie, M., Garin, J. & Attree, I. (2005). PscF is a major component of the *Pseudomonas aeruginosa* type III secretion needle. *FEMS Microbiol Lett* **253**, 95-101.

7. Kenjale, R., Wilson, J., Zenk, S. F., Saurya, S., Picking, W. L., Picking, W. D. & Blocker, A. (2005). The needle component of the type III secreton of *Shigella* regulates the activity of the secretion apparatus. *J Biol Chem* **280**, 42929-37.
8. Torruellas, J., Jackson, M. W., Pennock, J. W. & Plano, G. V. (2005). The *Yersinia pestis* type III secretion needle plays a role in the regulation of Yop secretion. *Mol Microbiol* **57**, 1719-33.
9. Blocker, A., Jouihri, N., Larquet, E., Gounon, P., Ebel, F., Parsot, C., Sansonetti, P. & Allaoui, A. (2001). Structure and composition of the *Shigella flexneri* "needle complex", a part of its type III secreton. *Mol Microbiol* **39**, 652-63.
10. Kubori, T., Sukhan, A., Aizawa, S. I. & Galan, J. E. (2000). Molecular characterization and assembly of the needle complex of the *Salmonella typhimurium* type III protein secretion system. *Proc Natl Acad Sci U S A* **97**, 10225-30.
11. Mota, L. J., Journet, L., Sorg, I., Agrain, C. & Cornelis, G. R. (2005). Bacterial injectisomes: needle length does matter. *Science* **307**, 1278.
12. Deane, J. E., Roversi, P., Cordes, F. S., Johnson, S., Kenjale, R., Daniell, S., Booy, F., Picking, W. D., Picking, W. L., Blocker, A. J. & Lea, S. M. (2006). Molecular model of a type III secretion system needle: Implications for host-cell sensing. *Proc Natl Acad Sci U S A* **103**, 12529-33.
13. Darboe, N., Kenjale, R., Picking, W. L., Picking, W. D. & Middaugh, C. R. (2006). Physical characterization of MxiH and PrgI, the needle component of the type III secretion apparatus from *Shigella* and *Salmonella*. *Protein Sci* **15**, 543-52.
14. Zhang, L., Wang, Y., Picking, W. L., Picking, W. D. & De Guzman, R. N. (2006). Solution structure of monomeric BsaL, the type III secretion needle protein of *Burkholderia pseudomallei*. *J Mol Biol* **359**, 322-30.
15. Dance, D. A. (1991). Melioidosis: the tip of the iceberg? *Clin Microbiol Rev* **4**, 52-60.
16. Currie, B. J., Fisher, D. A., Howard, D. M. & Burrow, J. N. (2000). Neurological melioidosis. *Acta Trop* **74**, 145-51.
17. Dance, D. A. (2000). Melioidosis as an emerging global problem. *Acta Trop* **74**, 115-9.
18. Chaowagul, W., Suputtamongkol, Y., Dance, D. A., Rajchanuvong, A., Pattara-arechachai, J. & White, N. J. (1993). Relapse in melioidosis: incidence and risk factors. *J Infect Dis* **168**, 1181-5.
19. Stevens, M. P., Wood, M. W., Taylor, L. A., Monaghan, P., Hawes, P., Jones, P. W., Wallis, T. S. & Galyov, E. E. (2002). An Inv/Mxi-Spa-like type III protein secretion system in *Burkholderia pseudomallei* modulates intracellular behaviour of the pathogen. *Mol Microbiol* **46**, 649-59.
20. Jones, A. L., Beveridge, T. J. & Woods, D. E. (1996). Intracellular survival of *Burkholderia pseudomallei*. *Infect Immun* **64**, 782-90.

21. Grzesiek, S. & Bax, A. (1993). The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115**, 12953-12954.
22. Tolman, J. R., Chung, J. & Prestegard, J. H. (2004). Pure-phase heteronuclear multiple-quantum spectroscopy using field gradient selection *J. Magn. Reson.* **98**, 462-467.
23. Grzesiek, S., Dobeli, H., Gentz, R., Garotta, G., Labhardt, A. M. & Bax, A. (1992). 1H, 13C, and 15N NMR backbone assignments and secondary structure of human interferon-gamma. *Biochemistry* **31**, 8180-90.
24. Muhandiram, D. R. & Kay, L. E. (1994). Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. ser. B* **103**, 203-216.
25. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989). Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry* **28**, 6150-6.
26. Fesik, S. W. & Zuiderweg, E. R. P. (1988). Heteronuclear three-dimensional NMR spectroscopy. A strategy for the simplification of homonuclear two-dimensional NMR spectra. *J. Magn. Reson.* **78**, 588-593.
27. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-93.
28. Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* **278**, 313-52.
29. Wishart, D. S. & Nip, A. M. (1998). Protein chemical shift analysis: a practical guide. *Biochem Cell Biol* **76**, 153-63.
30. Guntert, P. (2004). Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**, 353-78.
31. Dames, S. A., Martinez-Yamout, M., De Guzman, R. N., Dyson, H. J. & Wright, P. E. (2002). Structural basis for Hif-1 alpha /CBP recognition in the cellular hypoxic response. *Proc Natl Acad Sci U S A* **99**, 5271-6.
32. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* **8**, 477-86.
33. Koradi, R., Billeter, M. & Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* **14**, 51-5, 29-32.
34. Delano, W. L. "The Pymol Molecular Graphics System." Delano Scientific LLC, San Carlos, CA, USA. http://www.pymol.org/.
35. Wuthrich, K. (1986). NMR of Proteins and Nucleic Acids, Wiley-Interscience, New York, NY.

36. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J Mol Biol* **233**, 123-38.
37. Phan, J., Austin, B. P. & Waugh, D. S. (2005). Crystal structure of the *Yersinia* type III secretion protein YscE. *Protein Sci* **14**, 2759-63.
38. Blocker, A., Komoriya, K. & Aizawa, S. (2003). Type III secretion systems and bacterial flagella: insights into their function from structural similarities. *Proc Natl Acad Sci U S A* **100**, 3027-30.
39. Quinaud, M., Chabert, J., Faudry, E., Neumann, E., Lemaire, D., Pastor, A., Elsen, S., Dessen, A. & Attree, I. (2005). The PscE-PscF-PscG complex controls type III secretion needle biogenesis in *Pseudomonas aeruginosa*. *J Biol Chem* **280**, 36293-300.
40. Samatey, F. A., Matsunami, H., Imada, K., Nagashima, S., Shaikh, T. R., Thomas, D. R., Chen, J. Z., Derosier, D. J., Kitao, A. & Namba, K. (2004). Structure of the bacterial flagellar hook and implication for the molecular universal joint mechanism. *Nature* **431**, 1062-8.

SUPPLEMENTAL MATERIALS

Table 2-S1. Details of NMR experiments used for BsaL$^{C\Delta5}$.

NMR Experiment	Bruker Pulse Sequence	Experimental Details	Time*
2D HSQC	hsqcfpf3gpphwg	8 NS* 2048×128 (N) pts	20 min
2D HMQC	blhmqcnoesy2d	8 NS 2048×256(C) pts	1h 30min
3D HNCA	hncagpwg3d	16 NS 2048×48(N)×70(C) pts	18h
3D CBCA(CO)NH	cbcaconhgpwg3d	16 NS 2048×48(N)×100 (C) pts	1d
3D HNCO	hncogpwg3d	16 NS 2048×48(N)×44 (C) pts	1d 3h
3D HSQC-NOESY	noesyhsqcf3gp193d	16 NS 2048×44(N)×256 (H) pts	2d 17h
3D HMQC-NOESY	blhmqcnoesy3d	8 NS 2048×64(C)×200(H) pts	2d
Heteronuclear ^{1}H-^{15}N NOE	rdhsqcnoef3gpsi	24 NS 2048×360 (N) pts	19h

NS*: number of scans
Time*: total acquisition time

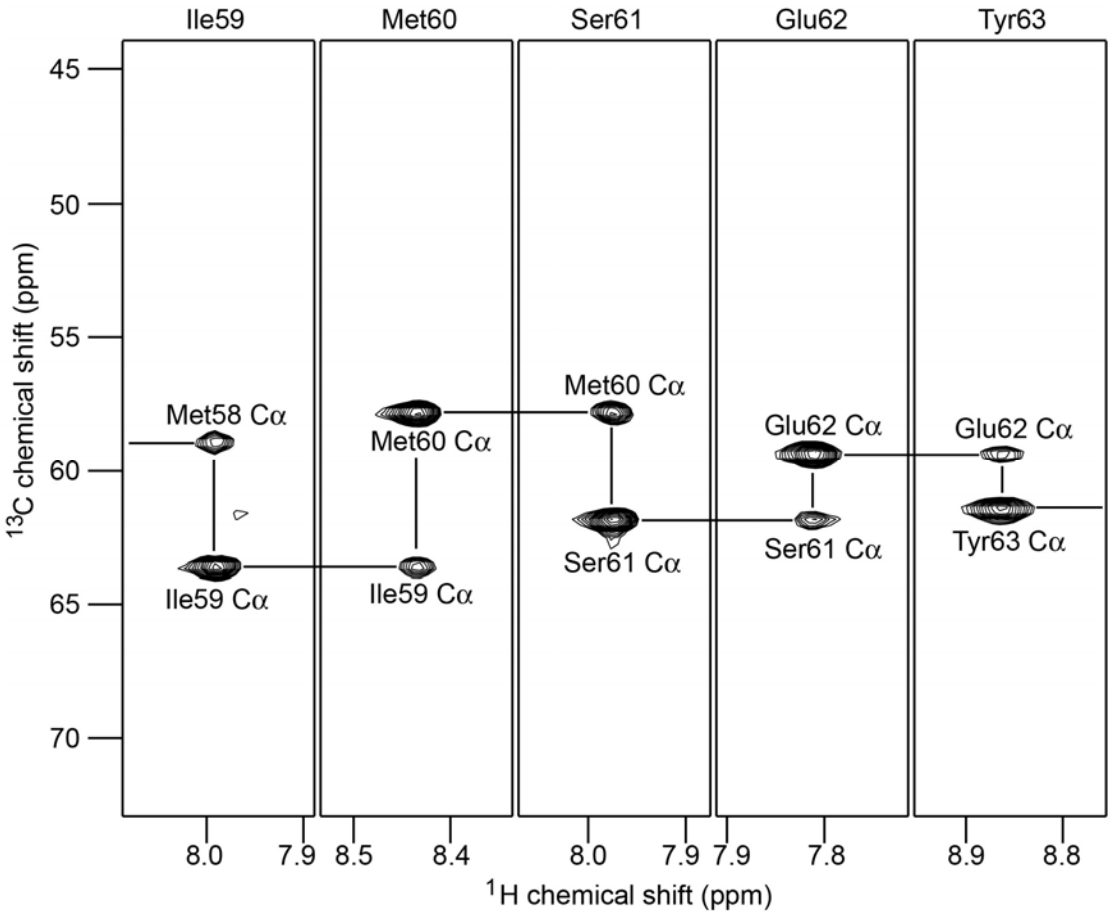

Figure 2-S1. Portion of the 3D HNCA strips of ^{13}C, ^{15}N-labeled BsaL$^{C\Delta5}$, shown as ^{13}C, ^{1}H strips for Ile59-Tyr63 (the ^{15}N ppm is on the Z-axis). Each strip shows the C_α of the preceeding and current residue and the amide (^{1}H and ^{15}N) resonances of the current residue, indicating a sequence specific connectivity between these strips. In theory, the 3D HNCA dataset contains all the information needed in assigning all the backbone resonances of a protein. However, peak overlaps make assignments ambiguous; and a combination with the CBCA(CO)NH and ^{15}N-HSQC-NOESY experiments were used to complete the backbone assignments of BsaL$^{C\Delta5}$.

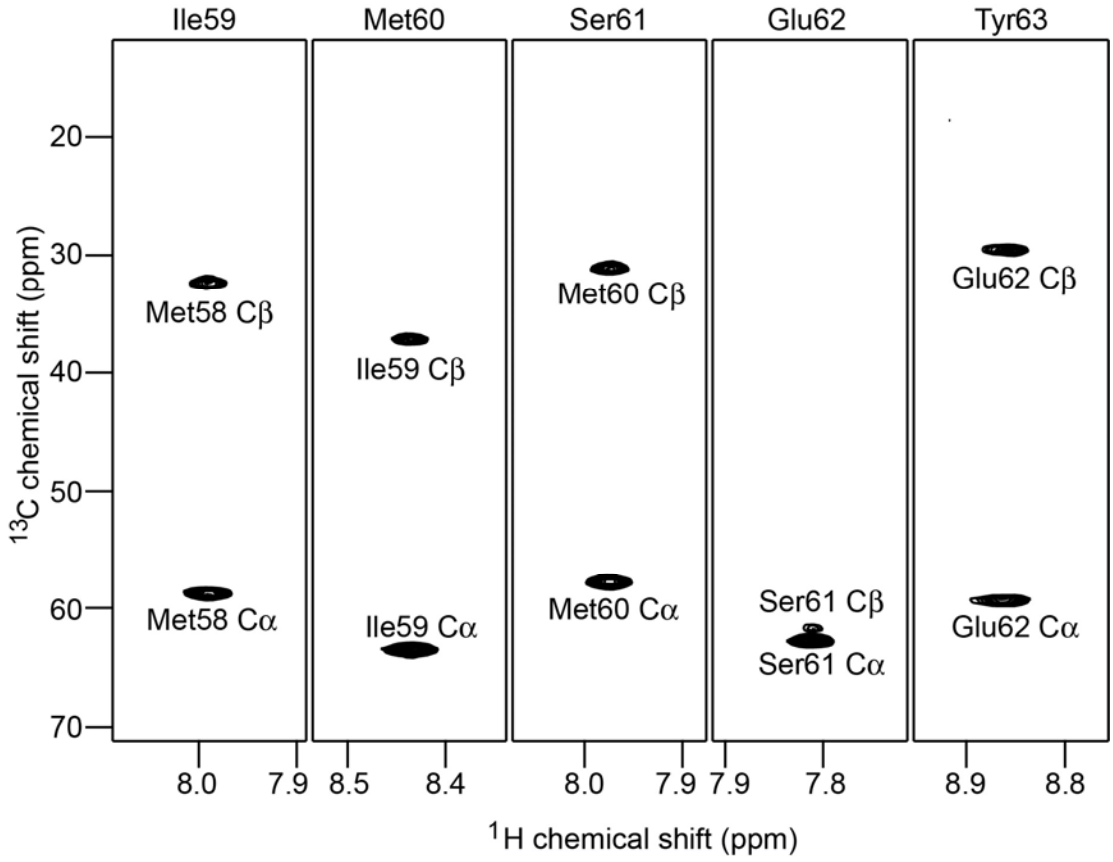

Figure 2-S2. Portion of the 3D CBCA(CO)NH strips of ^{13}C, ^{15}N-labeled BsaL$^{C\Delta5}$, shown as ^{13}C, ^{1}H strips for Ile59-Tyr63 (the ^{15}N ppm is on the Z-axis). Each strip shows the C_α and the C_β of the preceeding residue and the amide (^{1}H and ^{15}N) resonances of the current residue. 3D CBCA(CO)NH was collected to assist in the sequence specific assignment.

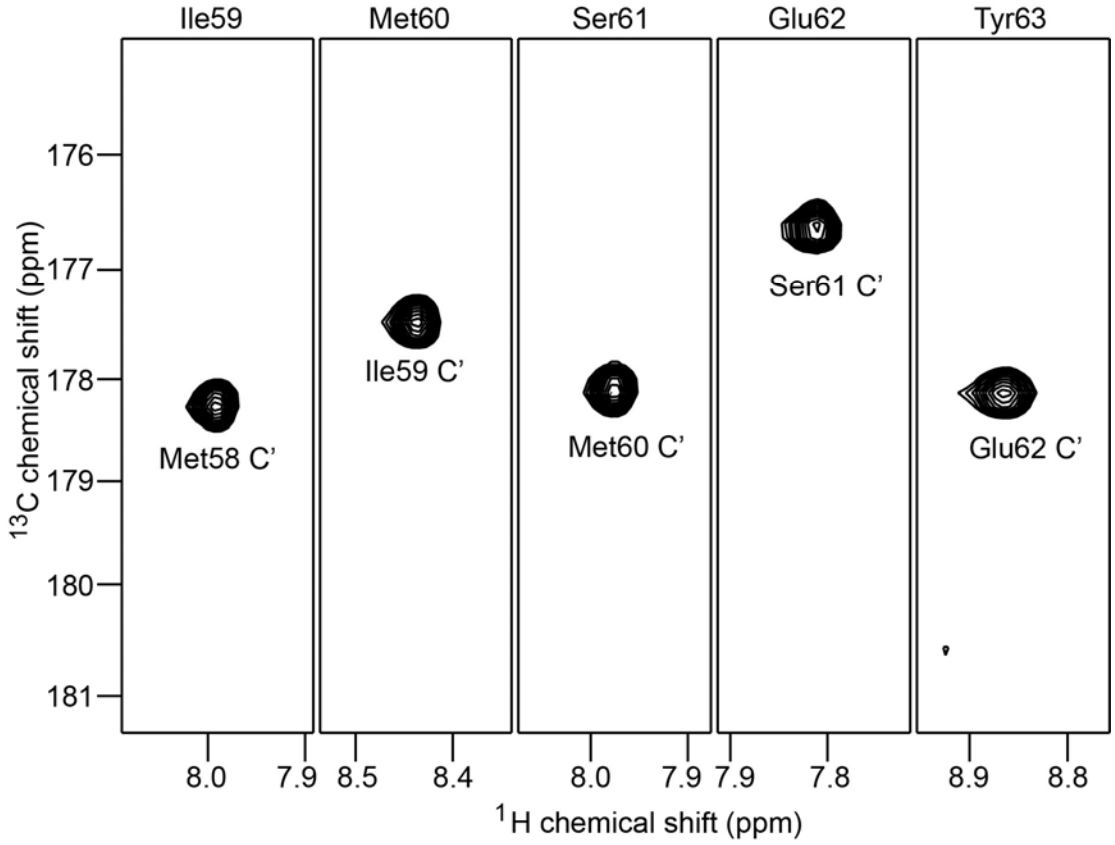

Figure 2-S3. Portion of the 3D HNCO strips of ^{13}C, ^{15}N-labeled $BsaL^{C\Delta5}$, shown as ^{13}C, ^{1}H strips for Ile59-Tyr63 (the ^{15}N ppm is on the Z-axis). Each strip shows the $^{13}C'$ (carbonyl) peak of the preceeding residue and the amide (^{1}H and ^{15}N) resonances of the current residue. HNCO dataset was collected to determine the carbonyl atom secondary chemical shifts of $BsaL^{C\Delta5}$, which is a good indicator of the presence of α-helix or β-strand in proteins.

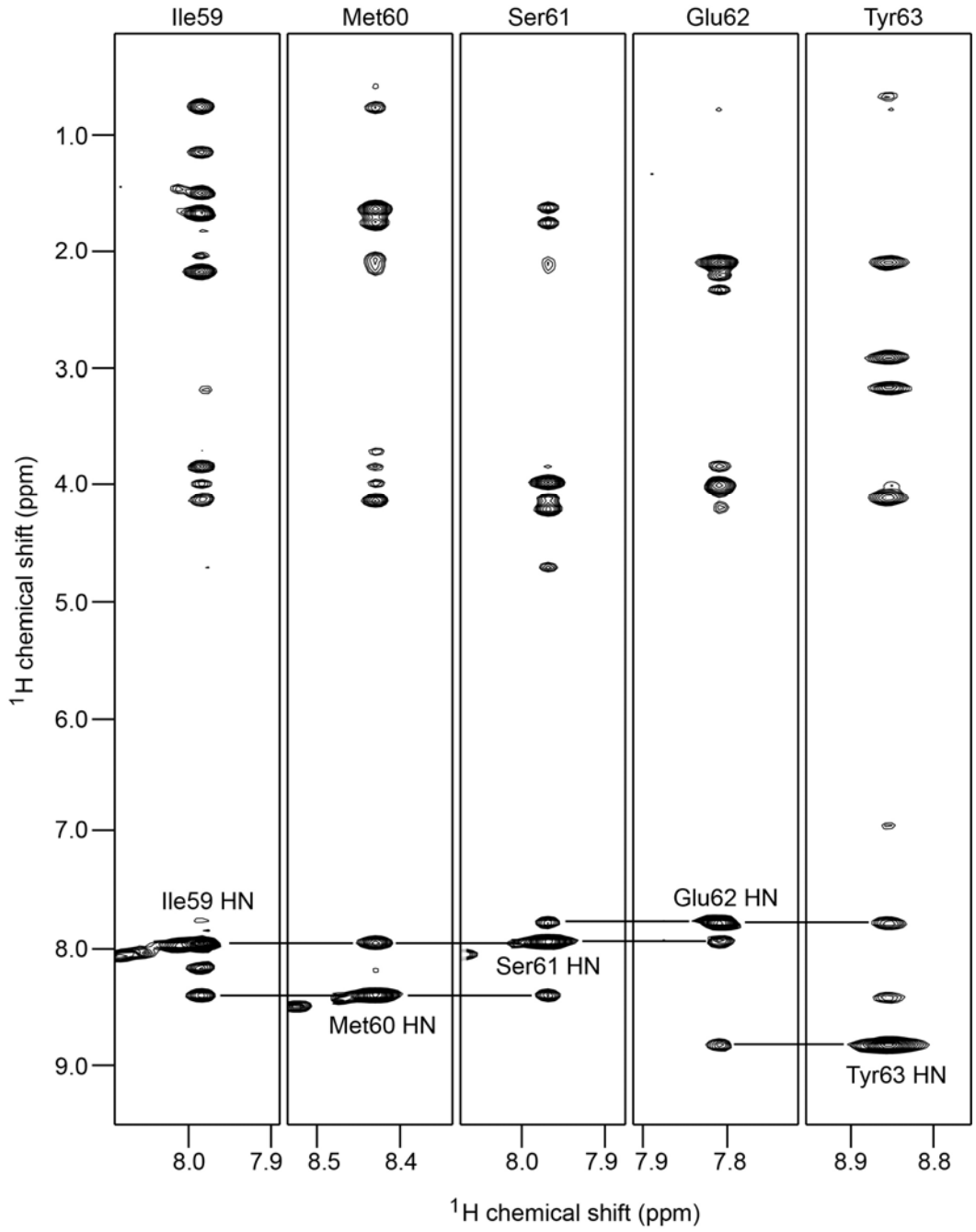

Figure 2-S4. Portion of the 3D ^{15}N-edited HSQC-NOESY strips of ^{15}N-labeled BsaL$^{C\Delta5}$, shown as ^{1}H, H^N strips for Ile59-Tyr63 (the ^{15}N ppm is on the Z-axis). Each strip shows all the backbone H_N NOE peaks including its own H_N and the H_Ns of the preceding and following residues, indicating connectivity between these residues which can facilitate in the backbone assignments of regions with α-helical secondary structure.

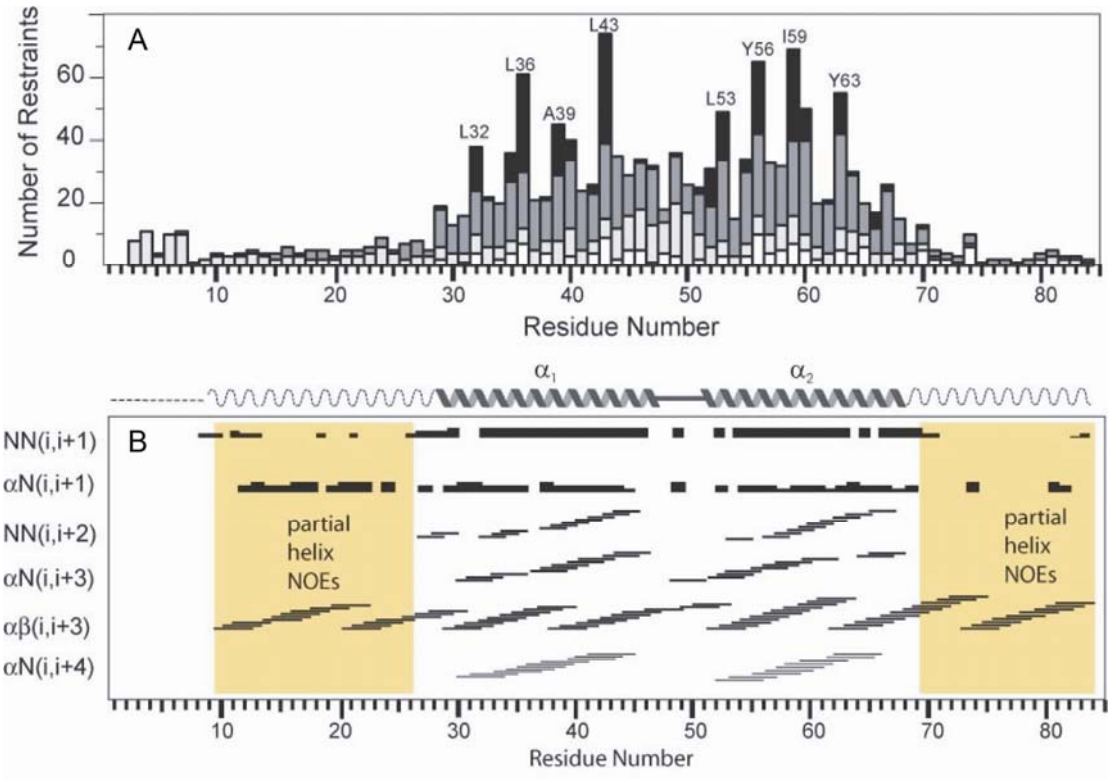

Figure 2-S5. The long-range NOE restraints (dark bars) are clustered in the in the α-helical regions of BsaL$^{C\Delta5}$ (V29-Q70). The flanking regions (A10-G28 and S71-I84) have medium range (gray bars) and sequential (white bars) NOEs. **(B)** The well defined two-helix bundle in the middle of BsaL (helix α1 and α2) have the characteristic α-helical NOE patterns. Both the N-terminal region (A10-G28, shaded) and C-terminal region (S71-I84) also contain medium-range α-helical NOEs.[14]

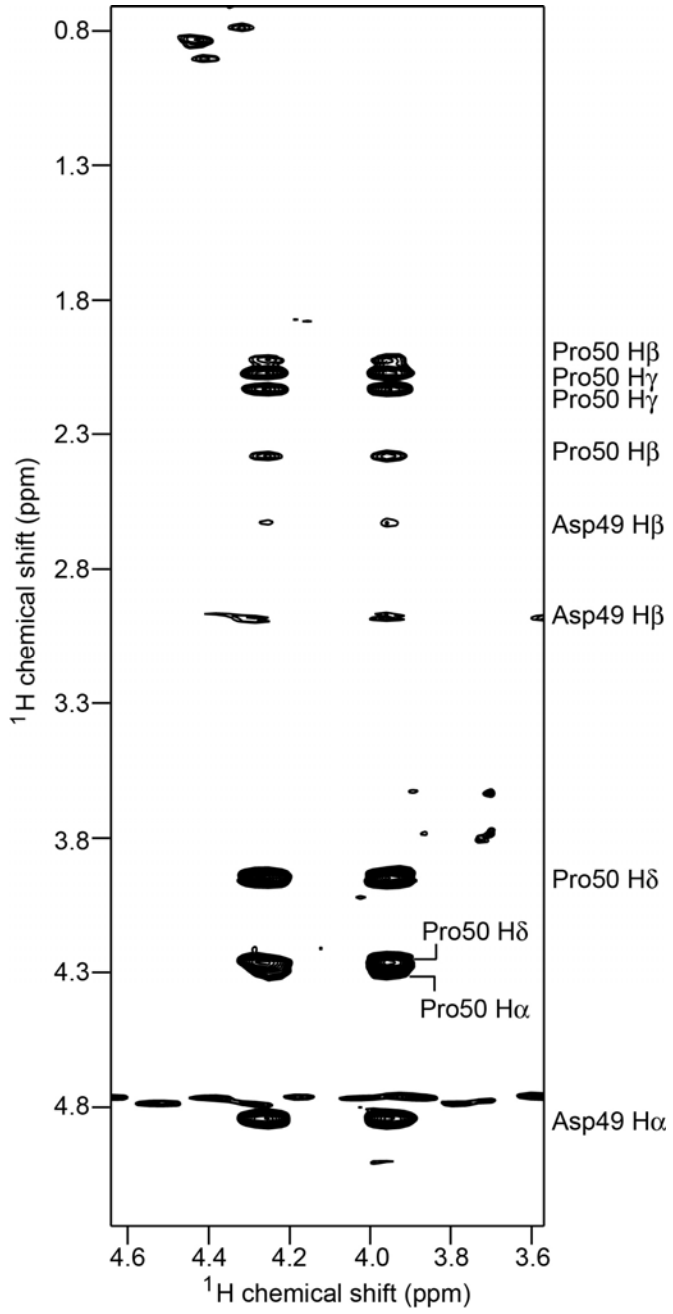

Figure 2-S6. Strips of the two H_δ protons of residue Pro50 in 3D ^{13}C-edited HMQC-NOESY of ^{13}C, ^{15}N-labeled BsaL$^{C\Delta5}$, showing the proline trans configuration. NOE peak intensity of previous residue Asp49 H_α was stronger than the Pro50 H_α. The Z-axis is at Pro50 C_δ chemical shift at 51.339 ppm.

Chapter 3: Solution Structure Determination of PrgI

INTRODUCTION

Salmonella typhimurium is a major intestinal pathogen associated with food poisoning and is estimated to cause 2 to 4 million cases of salmonellosis in the U.S. annually.[1] Salmonellosis contributes significantly to hospitalization and mortality among susceptible individuals such as the elderly, children, and people with compromised immune systems caused by AIDS and chemotherapy.[1]

Salmonella encodes two TTSSs on two S*almonella p*athogenicity *i*slands (SPI1 and SPI2).[2; 3; 4] SPI1 is necessary for invasion into the nonphagocytic cells of the intestinal epithelium.[5] *Salmonella* promotes its own uptake into nonphagocytic cells by subverting the normal function of several signaling pathways that control the formation of transient actin-rich membrane ruffles that engulf the infecting bacterium (Fig 1-2).[6] Inside the host cells, becteria will be allowed to reside and replicate in a privileged site, minimizing exposure to the immune system of the host.[6] SPI2 genes are expressed only after inside the host cells and are required for the later stage survival with macrophages and subsequent systemic infection.[7; 8; 9; 10]

S. typhimurium uses a type III secretion apparatus (TTSA) to inject host altering effector proteins into eukaryotic cells. On the surface of the *S. typhimurium*, PrgI monomers assemble into a needle-like structure, and electron microscopy revealed a tube-like structure with dimensions 130 Å × 800 Å (diameter *versus* length).[11] PrgI is homologous to BsaL (*Burkholderia*) and MxiH (*Shigella*). The first

reported atomic resolution structure of a needle monomer is the NMR structure of BsaL.[12] Other groups subsequently solved the crystal structure of the *S. flexneri* MxiH needle protein, which showed a disordered N-terminal region and a longer C-terminal helix (Fig. 3-1).[13] Nevertheless, the structured region of BsaL can be superimposed to MxiH with a 3.2 Å C_α root-mean-square deviation (RMSD) (Fig. 3-1). Later, the crystal structures of two other needle proteins, PscF[14] from *Pseudomonas aeruginosa* and YscF[15] from *Yersinia pestis*, were solved, but when bound to their two respective chaperones (Fig. 3-2). The structures of PscF[14] and YscF[15] are different from the structures of BsaL[12], MxiH[13], and PrgI[16], which is probably due to the methods used to prevent the self-polymerization of PscF and YscF. Deleting a small number of C-terminal residues does not lead to readily purifiable PscF and YscF monomers. The C-terminal 31 residues of PscF in the ternary complex (with its chaperones PscG and PscE) form a short ordered region and a short helix similar to the C-terminal helix of MxiH, leaving the rest of the protein disordered (Fig. 3-2). Similarly, residues 50–87 of YscF (with its chaperones YscG and YscE) form an α-helical hairpin with a five-residue linker from Ile64 to Asn68 and the first 49 residues displayed no electron density (Fig. 3-2). In both ternary complexes, more than half of PscF and YscF N-terminal residues, including the region equivalent to the PxxP motif, are missing or disordered in the crystals. The structures of PscF and YscF show how bacteria prevent the premature self-polymerization of needle monomers while inside the bacterial cytoplasm; however, no chaperones have been reported for BsaL, MxiH, or PrgI. Questions about whether these dramatic structural differences between PscF and

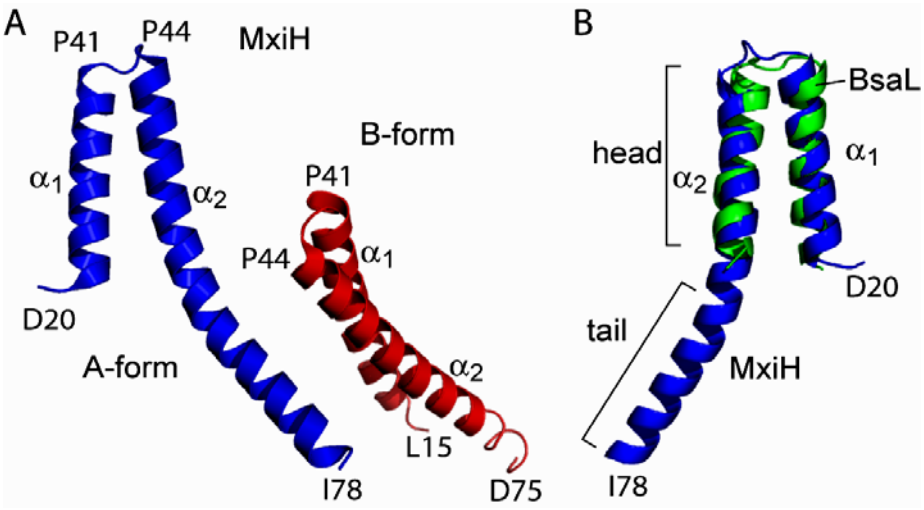

Figure 3-1. (**A**) Two different MxiH$^{C\Delta5}$ forms (form A and form B) were observed in the crystal packing (PDB 2CA5). Two forms differ on second helix α_2 and the boundary of the two helices. (**B**) The core domain two helices of MxiH$^{C\Delta5}$ form A superimposed with BsaL$^{C\Delta5}$ yields a 3.2 Å RMSD over C_α.

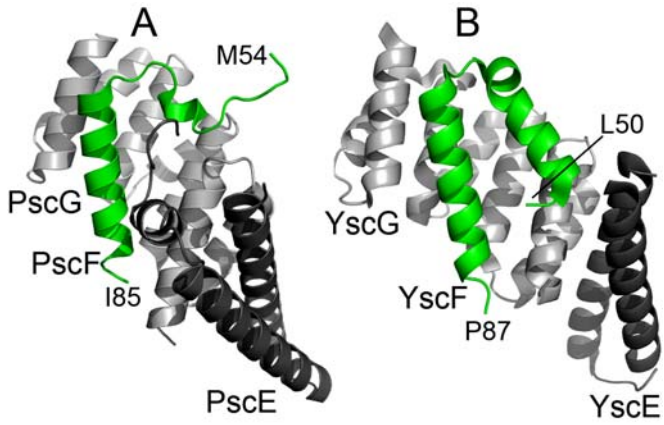

Figure 3-2. Needle proteins PscF (**A**) and YscF (**B**) together with their respective chaperones form ternary complexes. Only the structured part of the needle proteins PscF (M54-I85) and YscF (L50-P87) are shown in green, with the chaperones PscG/PscE and YscG/YscE colored in light grey and black, respectively.[17]

YscF with BsaL, MxiH and PrgI will lead to different needle packing remain unknown.

Despite sequence conservation, needle proteins have widely different thermal stabilities, with melting temperatures of 37°C for PrgI, 42°C for MxiH, and 54°C for BsaL[18]. In addition, PrgI does not complement a *S. flexneri mxiH* null mutant, implying that the needle proteins of *S. typhimurium* and *S. flexneri* are not interchangeable despite 63% sequence identity between PrgI and MxiH. These observations suggest there might be distinct structural features of PrgI, which makes the atomic structure determination of PrgI necessary for understanding its biological function.

MATERIALS AND METHODS

Expression and Purification of PrgI$^{C\Delta5}$

The plasmid expressing recombinant *S. typhimuriumi* PrgI lacking five C-terminal residues and possessing a His_6 C-terminal tag, which was obtained from Dr. William D. Picking (University of Kansas), was subcloned in the NdeI/XhoI sites of pET-22b vector (Novagen) as described.[18] PrgI$^{C\Delta5}$ is similar to the BsaL$^{C\Delta5}$ and MxiH$^{C\Delta5}$ constructs used in previous structure determinations where the last five residues were deleted and replaced with a C-terminal His_6-tag.[12; 13] Uniformly ^{13}C,^{15}N- and ^{15}N-labeled PrgI$^{C\Delta5}$ were obtained by growing *Escherichia coli* BL21(DE3) with the plasmid pET-22b-*prgI*$^{C\Delta5}$ in 1 liter of minimal media supplemented with 2 g/L of [^{13}C]glucose and/or 1 g/L of [^{15}N]ammonium chloride.

Cells were grown at 37°C, induced with 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside) at A_{600} 0.8, and expressed overnight at 15°C to a final A_{600} of 2.4. Cells were centrifuged, resuspended in 40 ml of buffer A (20 mM Tris (pH 7.9), 500 mM NaCl) with 5 mM imidazole, and sonicated. Cellular debris was removed by centrifugation and the supernatant was loaded on a 5 ml Ni^{2+}-affinity column (Sigma) that was previously treated with 35 ml of water, 35 ml of 50 mM $NiSO_4$, and 35 ml of buffer A with 5 mM imidazole. The Ni^{2+}-column was washed with 35 ml of buffer A with 60 mM imidazole and eluted with 15 ml of buffer A with 1 M imidazole in 1 ml fractions. Fractions were analyzed by SDS-PAGE gel and high concentration fractions were pooled. Protein samples were then dialyzed in NMR buffer (10 mM sodium phosphate pH 6.0, 10 mM NaCl) and protein concentration was estimated by UV absorbance at 280 nm.

NMR Spectroscopy of PrgI$^{C\Delta5}$

NMR data were acquired at 25°C using a 0.8 mM protein sample on a Bruker Avance 800 MHz spectrometer, processed with NMRPipe,[19] and analyzed using NMRView.[20] Resonance assignments were obtained from 2D ^{1}H-^{15}N HSQC[21] and ^{1}H-^{13}C HMQC[22] and the following 3D datasets: HNCA[23], HNCO[24], CBCA(CO)NH[23], HNCACB[25] and HBHACONH[26] (Table 3-S1). Water suppression was achieved using flip-back pulses and pulsed field gradients. The α-helices were identified from the C_α, C_β, and H_α secondary chemical shifts.[27] The ^{1}H, ^{13}C, ^{15}N chemical shifts were referenced indirectly using 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). Inter-

proton distance restraints were obtained from 3D ^{15}N-edited NOESY-HSQC (t_{mix} = 120 ms) and ^{13}C-edited HMQC-NOESY (t_{mix} = 120 ms) (Table 3-S1).

Salmonella Invasion Assay

The functional *Salmonella* invasion assay was done using the gentamycin protection invasion assay with cultured human epithelial cells. We followed a published protocol to examine the ability of *S. typhimurium* to invade the cultured human epithelial cell line Henle 407.[28] Henle 407 cells were obtained from the American Type Culture Collection (ATCC # CCL-6). *S. typhimurium* wild-type (SL1344) and *prgI* null mutant (JK17)[29] strains were obtained from Dr Bradley Jones (University of Iowa). Briefly, Henle 407 cells were grown in Eagle's modified minimum essential medium with 10% (v/v) newborn calf serum in 5% (v/v) CO_2. All *prgI* mutant plasmids (ampicillin and trimethoprim resistence) were transformed into *S. typhimurium* JK17 (ampicillin and kanamycin resistence) [16; 29] by electroporation. The bacteria were cultured fresh each time and the bacterial plates were streaked 2 days before the experiment. The following day, bacteria were grown in 10 ml LB media containing the appropriate antibiotics (kanamycin, ampicillin, and trimethoprim) in standing cultures (without shaking) at 37°C. When bacterial density reached A_{600}~0.5, 10~20 µl bacterial suspension was incubated with Henle 407 cells for 60 min to allow invasion before removal by aspiration. The Henle 407 cells were washed with DMEM media containing 50 µg/ml of gentamycin 3 times and further incubated in the gentamycin containing DMEM media for 90 min. The Henle 407

cells were then washed with DMEM (Dulbecco's Modified Eagle Medium) media and lyzed with 0.9% (w/v) deoxycholate to free the entrapped bacteria. The number of bacterial colonies, which correlates with invasiveness, were estimated by serial dilution and plating. All invasion assays for *prgI* point mutants together with *S. typhimurium* wide-type (SL1344) were done in triplicates.

RESULTS

NMR Structure Determination of PrgI$^{C\Delta5}$

The last five residues of PrgI were deleted to obtain monomeric PrgI$^{C\Delta5}$, similar to the approach used in the structure determinations of BsaL$^{C\Delta5}$ and MxiH$^{C\Delta5}$. Analysis of NMR datasets (Fig. 3-S1-S5) yielded near complete backbone and side chain assignments for PrgI$^{C\Delta5}$. We assigned 69 (out of 79) backbone amides (Fig. 3-3); however, the amides of Ala2, Thr3, Ala35, Ala36, Leu43 and Leu44 are in rapid exchange with the solvent and could not be assigned. The C_α, C_β, and H_α secondary chemical shifts indicated the presence of three helices: Val20–Ala35, Pro41–Ser62, and Asp70–Glu77 (Fig. 3-4). Assignment of 3D nuclear Overhauser effect spectroscopy (NOESY) datasets yielded 1270 interproton distance restraints, or an average of 15 NOEs per residue (Table 3-1). Most of the nuclear Overhauser effects (NOEs) are in the helical regions and all the long-range NOEs are from the conserved hydrophobic residues in the central region of PrgI (Fig. 3-S6). Structure calculations using NMR-derived restraints (Table 3-1) and CYANA[30] followed by refinement in AMBER[31] yielded well converged NMR structures of PrgI$^{C\Delta5}$ (Fig. 3-5). The NMR

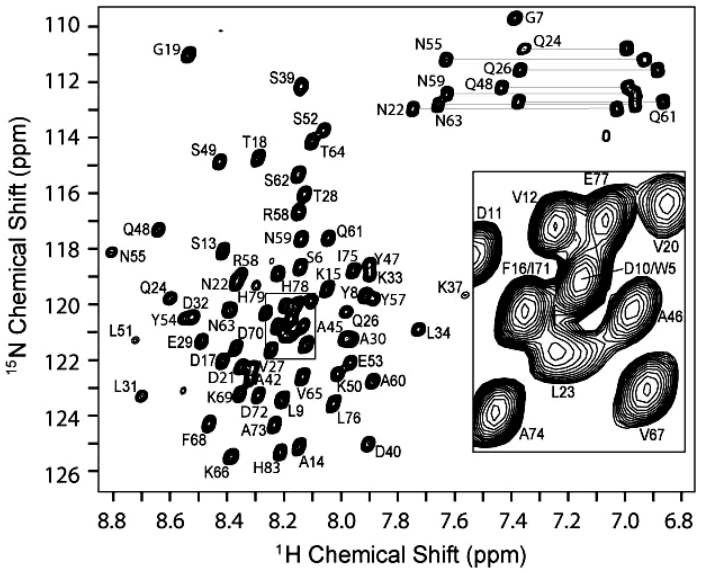

Figure 3-3. Assigned ^{1}H-^{15}N HSQC spectrum of PrgI$^{C\Delta5}$. The inset shows an expansion of the crowded region of the spectrum. The side-chain peak of Trp5 (not shown) is at 10.16 ^{1}H ppm and 129.89 ^{15}N ppm.[16]

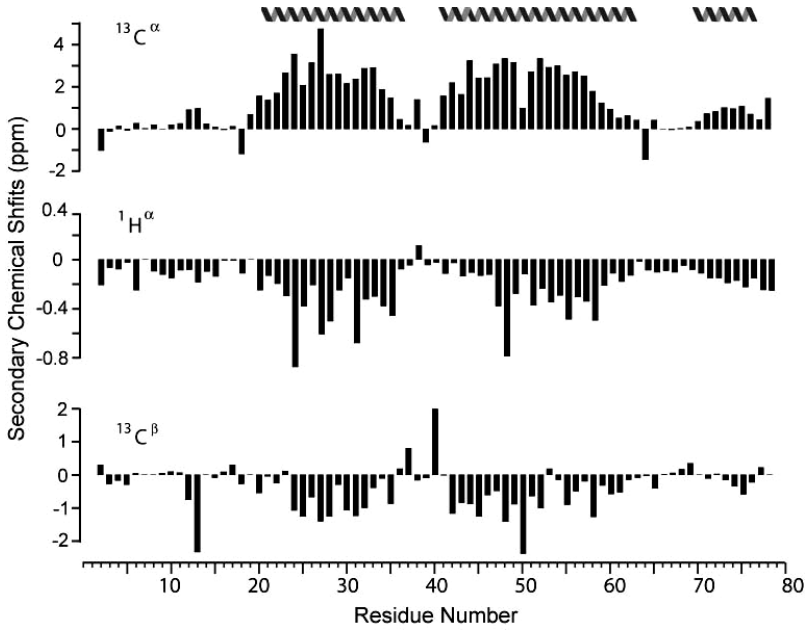

Figure 3-4. Secondary chemical shifts obtained by subtracting the random coil $^{13}C_{\alpha}$, $^{1}H_{\alpha}$ and $^{13}C_{\beta}$ chemical shift values from those observed in the protein indicate the approximate locations of the alpha-helical regions in PrgI$^{C\Delta5}$.[16]

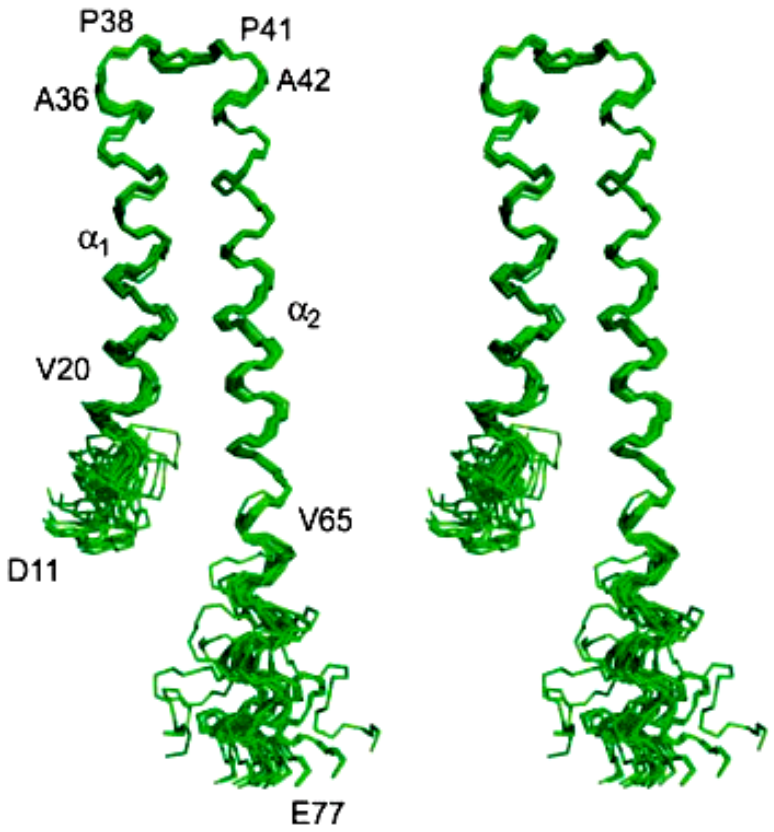

Figure 3-5. Stereo view of the superposition of 20 lowest energy structures of $PrgI^{C\Delta5}$, showing only the backbone N, C_α, and C′ atoms. Only the region Val20–Val65 can be superimposed into an ensemble of NMR structures, forming the core domain of $PrgI^{C\Delta5}$; however, the regions flanking this core domain, from Trp5–Gly19 and Lys66–Ile75, retained α-helical conformation in each of the 20 NMR structures.[16]

Table 3-1. Number of restraints and structural statistics for 20 NMR structures of $PrgI^{C\Delta5.16}$.

Total unambiguous distance restraints	1270
Intraresidue (*i,i*)	283
Sequential (*i, i*+1)	344
Medium range (2<= | *i–j* | <=4)	469
Long range (| *i–j* | >4)	174
Total dihedral angle restraints	101
Phi	68
Psi	33
RMS difference from mean structure (V20–V65)	
Backbone atoms of helices	0.26
All heavy atoms of helices	0.60
Violation analysis	
Maximum distance violation (Å)	0.3
Maximum dihedral angle violation (deg.)	4.4
Energies	
Mean AMBER energy (kcal mol^{-1})	−2464
Mean restraint energy (kcal mol^{-1})	114
Deviation from idealized geometry	
Bond lengths (Å)	0.00983±0.00005
Bond angles (deg.)	2.79±0.01
Ramachandran plot (%)	
Most favorable region	85.1
Additionally allowed regions	14.9
Generously allowed regions	0
Disallowed regions	0

structures yielded good structural statistics (Table 3-1), and over 85% of the residues are in the most favorable region of the Ramachandran plot.

Central Region Forms a Two-helix Bundle

Only the central region (Val20–Val65) of $PrgI^{C\Delta5}$ converged into a single family of NMR structures and formed an ordered two-helix bundle (Fig. 3-5). The regions flanking this two-helix bundle (Pro4–Gly19 and Lys66–Ile75) lacked long-range NOEs and did not converge; however, each of the 20 NMR structures showed a helical conformation for Pro4–Gly19 and Lys66–Ile75 that is consistent with the observed NOE patterns (Fig. 3-S6). The well-defined two-helix bundle is composed of the PxxP loop (Pro38–Pro41) and parts of helix α_1 (Val20–Ala36) and α_2 (Ala42–Val65). The helix α_1–α_2 interface (Fig. 3-6) is stabilized by conserved hydrophobic contacts of residues on helix α_1 (Val20, Leu23, Val27, Ala30, Leu34) and helix α_2 (Leu43, Leu44, Tyr47, Tyr54, Leu56). Val20 contacts the methylene groups of Gln24 and Arg58 and the phenyl ring of Tyr54 (Fig. 3-6A). L23 packs against the methylene groups of Gln26, Lys50, Glu53, Tyr54 and Tyr57 (Fig. 3-6); and Val27 is in a hydrophobic pocket formed by Gln24, Leu31, Tyr47, Leu51, and Tyr54 (Fig. 3-6B). The methyl group of the conserved Ala30 residue is in the helix α_1–α_2 interface but does not pack directly with any helix α_2 residue (Fig. 3-6A), suggesting that a small amino acid is required in this position. Finally, Leu34 packs with helix residues Leu31, Leu44, and Tyr47 and the PxxP loop residue Pro38 (Fig. 3-6B).

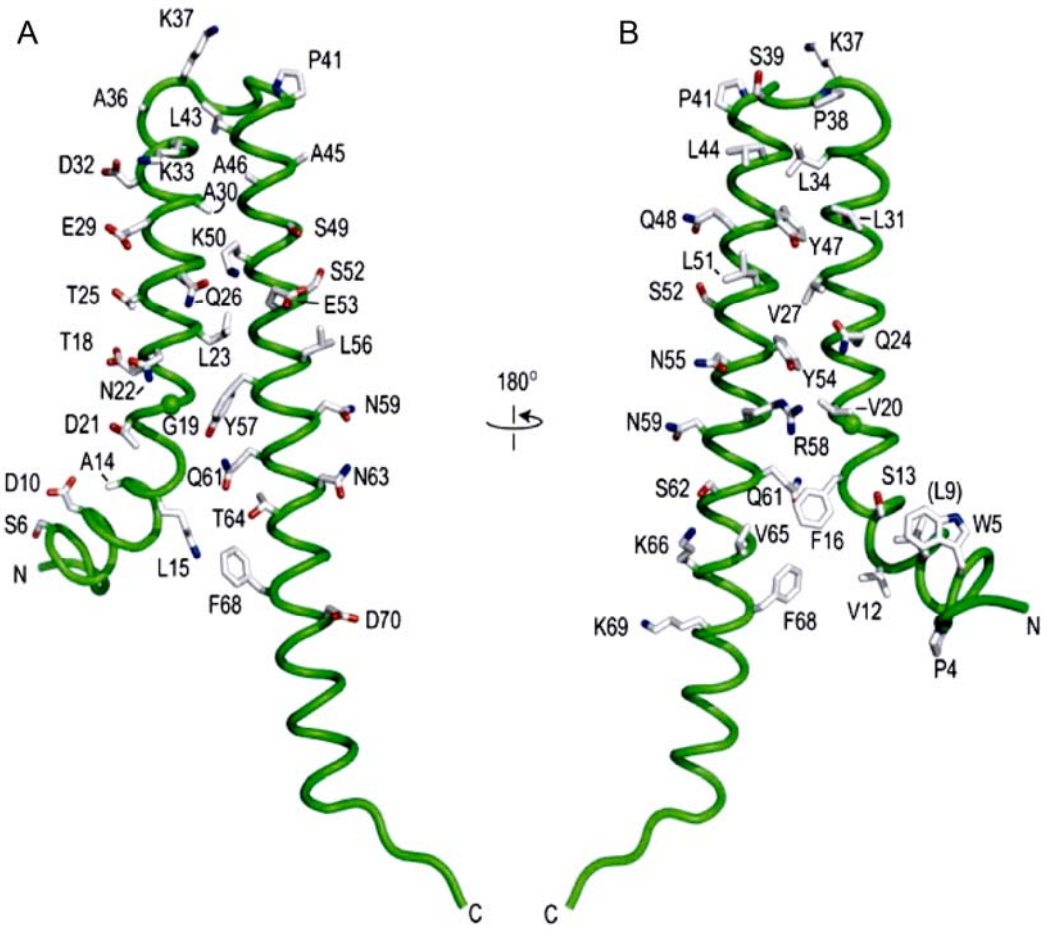

Figure 3-6. PrgI$^{C\Delta5}$ forms a two-helix bundle stabilized by hydrophobic contacts at the helix α_1–α_2 interface. The side-chains of the hydrophobic and polar residues are shown; **A** is oriented by a 180 rotation from **B**. Atoms are colored as: red, oxygen; blue, nitrogen; gray, carbon.[16]

The Surface of the Two-helix Bundle Is Polar

The surface of the two-helix bundle of PrgI is polar due to residues on helix α_1 (Asn22, Asp21, Gln24, Thr25, Gln26, Glu29, Lys33, and Asp32), helix α_2 (Gln48, Ser49, Lys50, Ser52, Glu53, Asn55, Arg58, Asn59, and Ser62), and the PxxP region (Lys27, Ser39, and Asp40) (Fig. 3-6). Most of these polar residues are pointed away from the two-helix bundle and therefore are not needed to stabilize the hydrophobic core of the two-helix bundle, however, several polar residues are identical (Gln48, Asn59, and Ser62) or conserved (Gln24, Glu53, Asn55, and Arg58) among needle proteins, suggesting that they are important for function other than stabilizing the core domain. Other residues that contribute to surface polarity are the conserved Tyr47 and Tyr54. While the Tyr47 and Tyr54 phenyl rings are required to stabilize the hydrophobic core, their hydroxyl groups are pointed towards the polar surface (Fig. 3-6B). Another polar moiety is the Gln26–Lys50 hydrogen bond in the middle of the two helix bundle of PrgI (Fig. 3-6A), which is in a position equivalent to the MxiH Glu29-Lys53 salt bridge.[13] BsaL, however, does not have any electrostatic interactions equivalent to that of PrgI Gln26-Lys50 or MxiH Glu29-Lys53.[13] Residues Gln24, Tyr54, Asn55, and Arg58 are clustered together on the surface of PrgI and form a polar patch at the base of the two-helix bundle (Fig. 3-6B).

Electrostatic Surfaces Are Different

The needle monomers are acidic proteins with theoretical pI's of 4.76 (PrgI and BsaL) and 4.47 (MxiH). Indeed, the electrostatic maps show large areas of

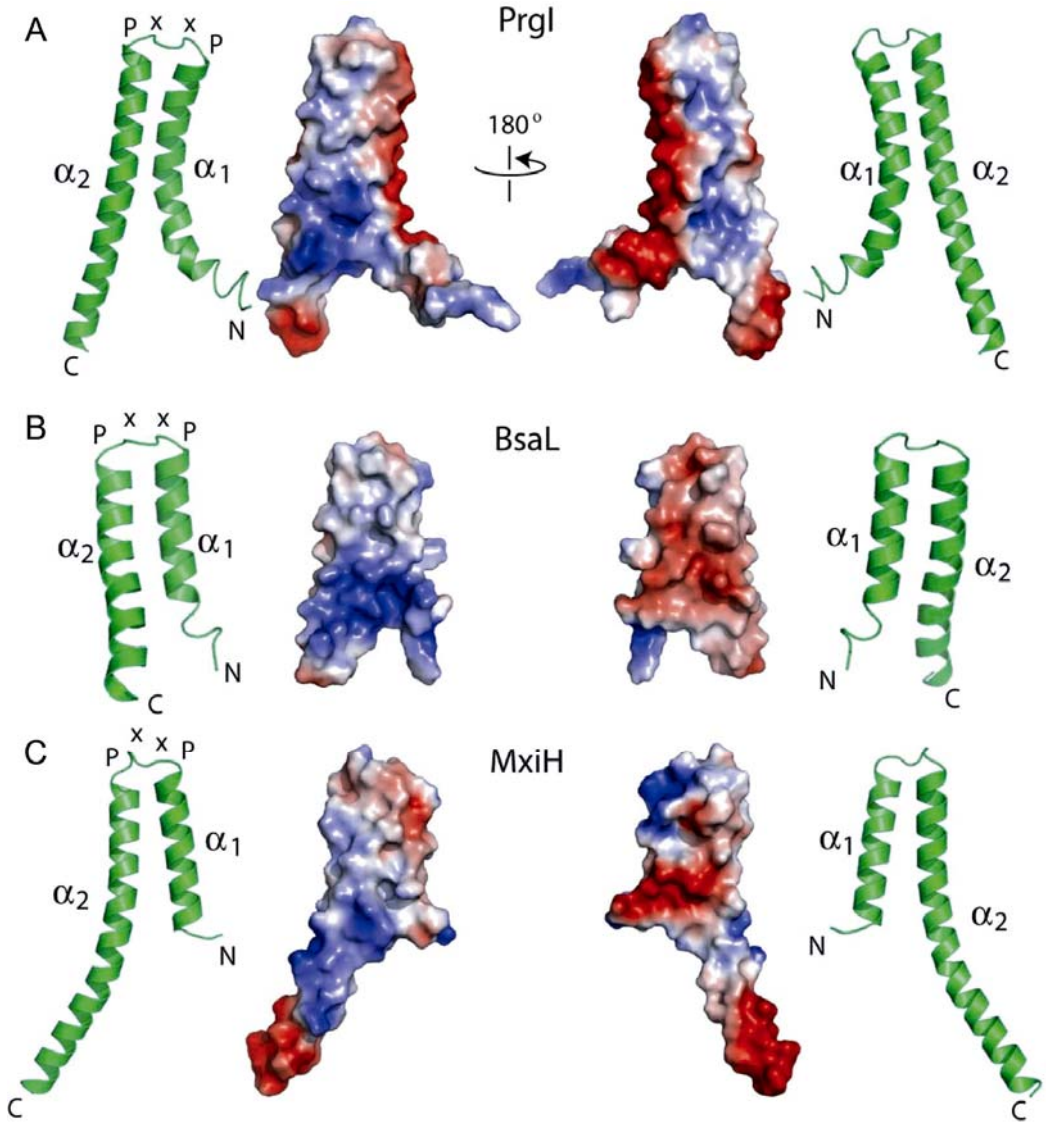

Figure 3-7. The electrostatic surface of (**A**) PrgI, (**B**) BsaL, and (**C**) MxiH are shown together with the ribbon structures showing the relative orientation of the two-helix bundle. The structures of PrgI, BsaL, and MxiH are oriented is similar manner from top to bottom to allow comparison of the surfaces. The structures on the left are rotated by 180° along the *y*-axis from the right.[16]

negatively charged surfaces (Fig. 3-7). The locations of the negatively charged surfaces, however, are radically different among PrgI, BsaL, and MxiH. In PrgI, the largest negatively charged surface is located on the "side" of the two–helix bundle and spans almost the entire length of helix α_1, from the PxxP motif to the base of the two helix bundle (Fig. 3-7A). This is in sharp contrast to BsaL and MxiH where the negatively charged surfaces are on the "face" of the two-helix bundle (Fig. 3-7B and c). In BsaL, the negatively charged surface is more broadly distributed on the helix α_1–α_2 interface (Fig. 3-7B). Likewise in MxiH, a negatively charged surface potential is on the "face" of the two-helix bundle (Fig. 3-7C). In PrgI, residues mostly on helix α_1 (Ser6, Asp10, Ser13, Thr18, Asn19, Asp21, Thr25, Glu29, and Asp32) contribute to the largest negatively charged surface, whereas, in BsaL and MxiH, residues on helix α_1 and α_2 contribute to the negatively charged face. The tail of helix α_2 in PrgI and MxiH (Fig. 3-7A and 7C), which contains conserved aspartic acids (PrgI Asp70 and Asp72), also has a significant negatively charged surface.

Despite being highly acidic proteins, PrgI, BsaL, and MxiH contain contiguous regions of basic (or positive) electrostatic potentials (Fig. 3-7). These positively charged regions are located similarly among the needle proteins, residing near the base of the two-helix bundle and distal to the PxxP motifs (Fig. 3-7). In PrgI and BsaL, the positively charged regions at the base of the two-helix bundle are due to residues on both helix α_1 and α_2 (Fig. 3-7A and 7B). In MxiH, however, the positively charged surface is solely due to residues on helix α_2, starting from the base of the two-helix bundle and extending towards the C-terminal end of helix α_2 (Fig. 3-

7C). In PrgI, the positively charged surface is located ~90° away from the negative side (Fig. 3-7A), whereas in BsaL and MxiH, the positively charged surfaces are 180° away, or on the flip side of the negatively charged surfaces (Fig. 3-7B and 7C).

Mutagenesis and *Salmonella* Invasion Assay

We used mutagenesis and assayed the ability of *Salmonella* to invade a cultured human epithelial cell line to assess the role of specific PrgI residues with respect to invasiveness (Fig. 3-S7). Wild-type bacteria form functional needles and can invade and give rise to hundreds of colonies, whereas a *prgI*-null mutant strain is non-invasive because it lacks a functional PrgI protein (Fig. 3-8). A plasmid that introduces wild-type *prgI* restores the invasiveness of the *prgI* null mutant. We mutagenized Gln26 and Lys50 to probe the importance of the Gln26–Lys50 hydrogen bond found in the middle of the two-helix bundle. We hypothesize this interaction contributes, albeit weakly, to stabilize the two-helix bundle, and therefore plays a role in needle assembly and virulence. Indeed, K50D, which is expected to disrupt the Gln26–Lys50 hydrogen bond, was shown *S. typhimurium* noninvasive (Fig. 3-8). The CD data confirmed that K50D was shown unstructured by CD spectroscopy (Fig. 3-8 and 3-9), suggesting the electrostatic repulsion (from a basic residue K to an acidic residue D) prevented the proper folding of PrgI into the two-helix bundle structure. K50L cannot participate in polar interactions yet it has wild-type level of invasiveness, suggesting it is capable of forming functional needles, and that the Gln26-Lys50 hydrogen bond is not critical for needle assembly and virulence. Mutant

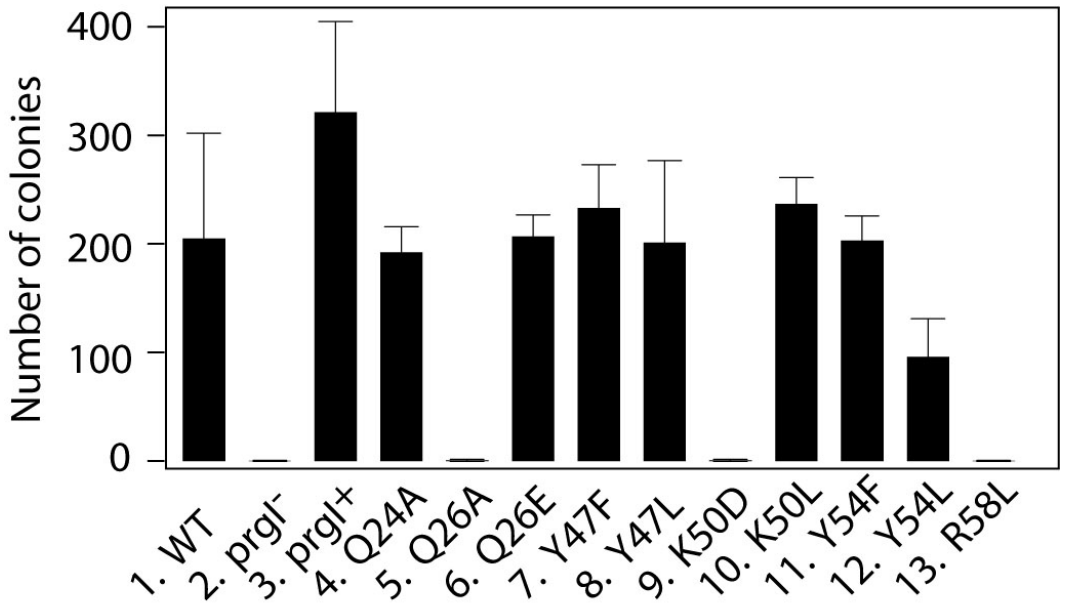

Figure 3-8. *Salmonella* invasion assay. The number of bacterial colonies (y-axis) correlates with the ability of *S. typhimurium* to invade a cultured human epithelial cell line (Henle 407). Wild-type (WT) bacterium is invasive, whereas a *S. typhimurium* with a *prgI* null mutation (*prgI*⁻) is non-invasive. A plasmid that overexpressed a functional PrgI (*prgI*$^{+}$) restored the invasiveness of the *prgI* null mutant strain. Point mutations were introduced in the Gln24, Gln26, Tyr47, Lys50, Tyr54 and Arg58 residues using the PrgI plasmid. The assays were done in triplicate.

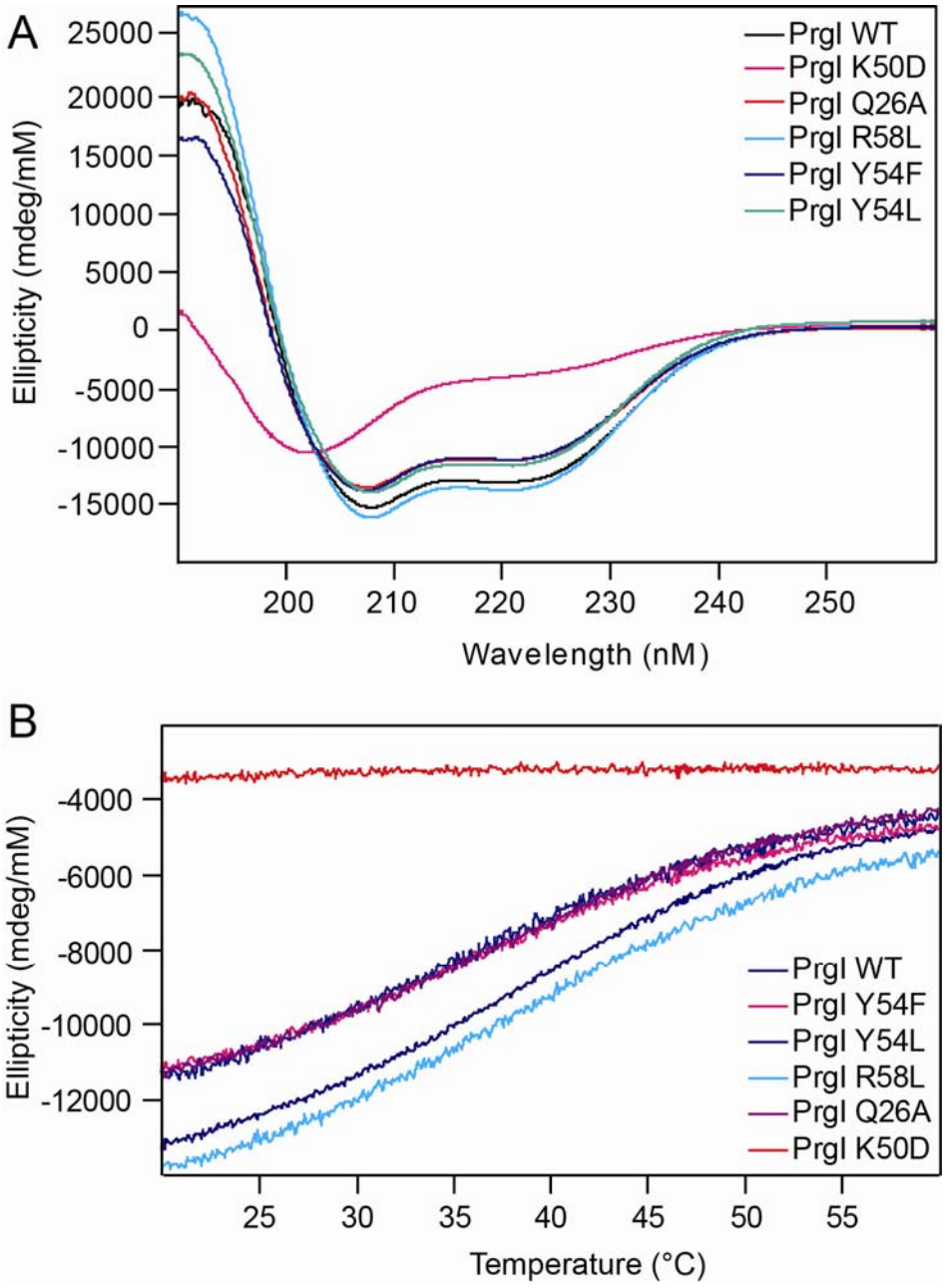

Figure 3-9. (**A**) CD spectra of PrgI$^{C\Delta5}$ wild type (WT) and point mutants (Q26A, Y54F, Y54L, R58L and K50D) showing the characteristic α-helical dips at 208 nm and 222 nm except for the K50D mutant. K50D shows a typical CD spectrum of a random coil protein, indicating K50D is unstructured. (**B**) CD thermal denaturation curves, monitored at 222 nm, of the wild type and PrgI point mutants. Except for K50D, all proteins show similar melting temperatures around 37°C.

Q26A, which is also expected to disrupt the hydrogen bond and was shown *S. typhimurium* noninvasive, however, was still able to form foled structure similar to wild-type PrgI indicating Q26 might be involved in the protein-protein interaction (Fig. 3-8 and 3-9). Q26E has wild-type invasiveness because glutamate can form a salt bridge with Lys50. Y54L, not Y54F, showed partially defect in *S. typhimurium* invasiveness, suggesting the aromatic ring might contribute to the protein-protein interaction. Finally, changing the electrostatic surface with the R58L mutation resulted to a non-invasive phenotype (Fig. 3-8), suggesting a function similar to residue Q26 and Y54.

DISCUSSION

TTSA needle proteins self-associate, but a C-terminal deletion was exploited to obtain monomeric $PrgI^{C\Delta5}$.[18] This is similar to the approach used in structure determinations for $BsaL^{C\Delta5}$ [12] and $MxiH^{C\Delta5}$ [13]. PrgI has 52% and 63% sequence identity with BsaL and MxiH, respectively, and the central region is more conserved compared to the N terminus (Fig. 2-1). $PrgI^{C\Delta5}$ (Fig. 3-3) and $BsaL^{C\Delta5}$ [12]gave excellent NMR spectra compared to $MxiH^{C\Delta5}$, allowing NMR structure determination. The central region of $PrgI^{C\Delta5}$ (Fig. 3-5), $BsaL^{C\Delta5}$,[12] and $MxiH^{C\Delta5}$ [13] form a core domain composed of two helices that are linked by a PxxP motif. Helix α_2 is longer than helix α_1, which is probably due to glycine and proline residues at the N termini, which act as helix breakers. The first 3, 8, and 19 residues of PrgI, Bsal, and MixH, respectively, are disordered. In PrgI and BsaL, the N and C-terminal

regions flanking the two-helix bundle have residual α-helical conformation and are more flexible. The first 19 residues are disordered in the MxiH$^{C\Delta5}$ crystal, however, a portion of this region might be in partial α-helical conformation when viewed by NMR as in BsaL and PrgI. The regions flanking the two-helix bundle do not fold back into the core structure and give the needle proteins an elongated shape instead of a globular structure.

The conserved hydrophobic contacts at the helix α_1–α_2 interface (Fig. 3-6) stabilize the two-helix bundle and are important in needle assembly and pathogenesis. More information is known about the effects of mutations on MxiH compared to PrgI and BsaL because Kenjale and co-workers[32] have characterized the phenotypes of over 35 MxiH point mutants (to alanine) with respect to needle assembly, secretion of effector proteins, and the ability of *S. flexneri* to invade cultured epithelial cells. In *S. flexneri*, MxiH L54A or Y57A mutation is defective in needle polymerization and secretion of effector proteins and thus is non-invasive.[32] The MxiH L54A or Y57A (equivalent to PrgI Leu51 and Tyr54) mutation is expected to decrease the hydrophobic contacts at the helix α_1–α_2 interface and destabilize the two-helix bundle. The proper folding of the two-helix bundle is thus essential in needle assembly.

The polar surface of the two-helix bundle is another common feature among needle proteins (Fig. 3-6). Conserved tyrosine residues contribute to surface polarity, however, the structures of PrgI, BsaL, and MxiH do not show a particular need for tyrosine residues in position Tyr47 and Tyr54; and phenylalanine would have maintained similar hydrophobic contacts and the overall fold of the two-helix bundle.

Yet, Tyr47 and Tyr54 are identical among PrgI, BsaL, and MxiH, suggesting the tyrosyl hydroxyl group is important for function. We hypothesize that the conserved tyrosine residues are important in protein–protein interactions with other needle monomers or with the tip complex.[33] Indeed, a MxiH Y50F point mutation, which is equivalent to PrgI Tyr47, disrupted the *in vivo* interaction between the MxiH needle and the tip complex and decreased the invasiveness of *S. flexneri* by 78%.[34] Another mutation affecting a polar residue, MxiH Q51A (equivalent to PrgI Gln48) showed normal needle assembly but decreased invasiveness.[32] Structurally, MxiH Gln51 is pointed away from the two–helix bundle and does not appear to contribute in stabilizing the core structure of the needle protein. However, MxiH Q51 is identical among the needle proteins (Fig. 2-1), suggesting that the decreased invasiveness of *S. flexneri* is probably due to defective protein–protein interaction. The polar surface of the needle protein is therefore important in pathogenesis.

Previous results showed that the last five residues are important in needle monomer–monomer interaction.[18; 32] Other than the last five residues, our data suggest that electrostatic interaction is also an important component of needle assembly. This is based on the presence of distinct areas of positively and negatively charged surface potentials on PrgI, MxiH, and BsaL (Fig. 3-7). Furthermore, this work shows that the most significant difference among these needle proteins is in the arrangement of their electrostatic surfaces (Fig. 3-7), suggesting that despite primary sequence conservation, protein–protein interaction may be different among these needle proteins. This could account for the observation that despite 63% sequence

identity between PrgI and MxiH, PrgI does not complement a *S. flexneri mxiH* null mutant (W.D. Picking, unpublished data). The needle proteins of *S. typhimurium* and *S. flexneri* are therefore not interchangeable. This also suggests that factors affecting electrostatic interactions, such as ionic strength, metal cations, and pH will affect needle assembly. The effect of salt concentration and cations on needle assembly have not been explored in detail; however, Marlovits *et al.*[35] showed *in vitro* that increasing the pH from pH 8 to pH 10.5 dissociated the needle, but not the basal structure of the *S. typhimurium* needle apparatus. We identified that a conserved tyrosine residue (equivalent to PrgI Tyr54) is involved in needle monomer–monomer contact by NMR chemical shift mapping of BsaL dimers (Fig. 3-S8 and S9). A possible scenario is that high pH changes the protonation of this Tyr54 hydroxyl group, which then interferes with needle monomer–monomer contact that leads to the dissociation of the needle.

The *Salmonella* invasion assay relies on the proper assembly of functional needles, therefore, it provides an indirect readout of the PrgI–PrgI interaction involved in needle assembly. We mutagenized the polar residues (Gln26, Lys50, Arg58) that contribute to the electrostatic surface of PrgI and used the *Salmonella* invasion assay to test their roles in needle assembly (Fig. 3-8). Mutagenesis of Gln26 and Lys50 suggest that the Gln26-Lys50 hydrogen bond is not critical for invasiveness, but the polarity of this region, either through Gln26 or Lys50, and the aliphatic moieties of Gln26 and Lys50 are required in needle assembly. Mutagenesis also suggests that the hydrophobic packing interaction of the aliphatic groups of

Gln26 and Lys50 are more important in the structure of the two-helix bundle than the hydrogen bonding interaction. Thus, Q26A and K50D, which have decreased aliphatic moieties, are non-invasive, whereas Q26E and K50L, which have aliphatic parts comparable to wild type, are invasive. The aliphatic part of Arg58, which is identical in PrgI, BsaL, and MxiH, participates in intramolecular hydrophobic contacts; however, its guanidine group is not involved in any polar interactions but contributes prominently to the positively charged surface of the two-helix bundle (Fig. 3-7). R58L mutation is non-invasive, suggesting that altering this surface polarity is deleterious in needle assembly.

Using X-ray fiber diffraction and electron microscopy, others showed that the *S. flexneri* needle is a helical structure formed by self-assembled needle monomers (Fig. 1-3).[36; 37] Combining the reported needle helical parameters and the crystal structure of the $MxiH^{C\Delta5}$ monomer, an atomic model of the *S. flexneri* needle was described at 16 Å resolution.[13] In their model, Deane *et al.* assumed that the N-terminal 19 residues, which were disordered in the $MxiH^{C\Delta5}$ crystal structure, formed a well-defined alpha helix that lines the inner channel (Fig. 3-10). They made this assumption because secondary structure predictions indicate the N-termini of needle proteins were α-helical.[32; 38] The major axis of the MxiH monomers was in line with the major axis of the needle. The orientation of this model is in agreement with the functionally- and structurally-related bacterial flagellar filament assembly.[13; 39; 40] The PxxP motif, which forms part of the head domain of MxiH, is pointed outwards on the needle. The disordered N-terminal region of MxiH is modeled to form a regular

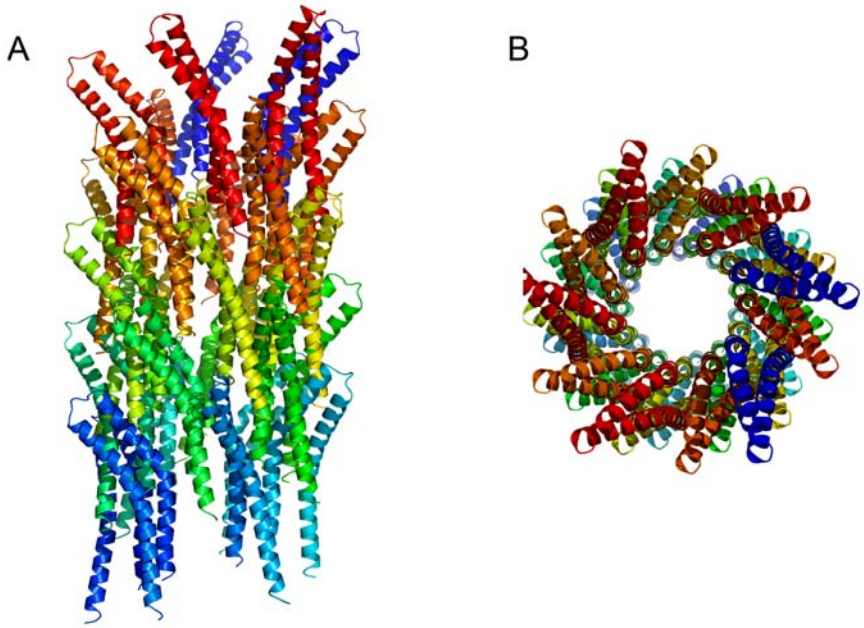

Figure 3-10. The crystal structure of MxiH monomer and electron microscopy of *Shigella* needles led to the first atomic model of a TTSA needle (PDB ID 2v6l). Shown are the (**A**) *side view* and (**B**) *end-on view* of the needle model.[17]

α-helix that lines the inner surface of the channel. The C-terminus of MxiH is buried within the needle wall and is in direct contact with three other MxiH monomers, thus showing how deletion of five C-terminal residues prevented the self-polymerization of MxiH.[32]

To examine the electrostatic contacts between needle monomers, Thenmalarchelvi Rathinavelan and Wonpil Im modeled the *S. flexneri* needle (Fig. 3-11A–D) using the $MxiH^{C\Delta5}$ crystal structure and the parameters reported in the *Shigella* needle model.[36; 37] The model was built with 28 subunits of the $MxiH^{C\Delta5}$ crystal structure using the CHARMM biomolecular simulation proGram.[41] The first 19 N-terminal residues and the last five C-terminal residues were also modeled to form regular α-helices. The N-terminal helix was also modeled to be inside the needle channel and parallel to the axis of the channel (the *Z*-axis). Each subunit was translated by 4.31 Å along the *Z*-axis and rotated by 64.3° around a 25 Å diameter of the needle channel. The model was energy minimized with a generalized Born implicit solvent model[42] in CHARMM to allow the side-chains to adopt low-energy conformations and resolve the steric clash between the residues at the binding interfaces. To model the *S. typhimurium* needle, the NMR structure of PrgI was superimposed on the backbone atoms (N, C_α, C′) of the MxiH needle and energy minimized.

Within the needle model, the two subunits with the largest surface contact were isolated and the electrostatic potentials of this dimer were calculated by solving the linearized Poisson–Boltzmann equation using the PBEQ module[43; 44] in

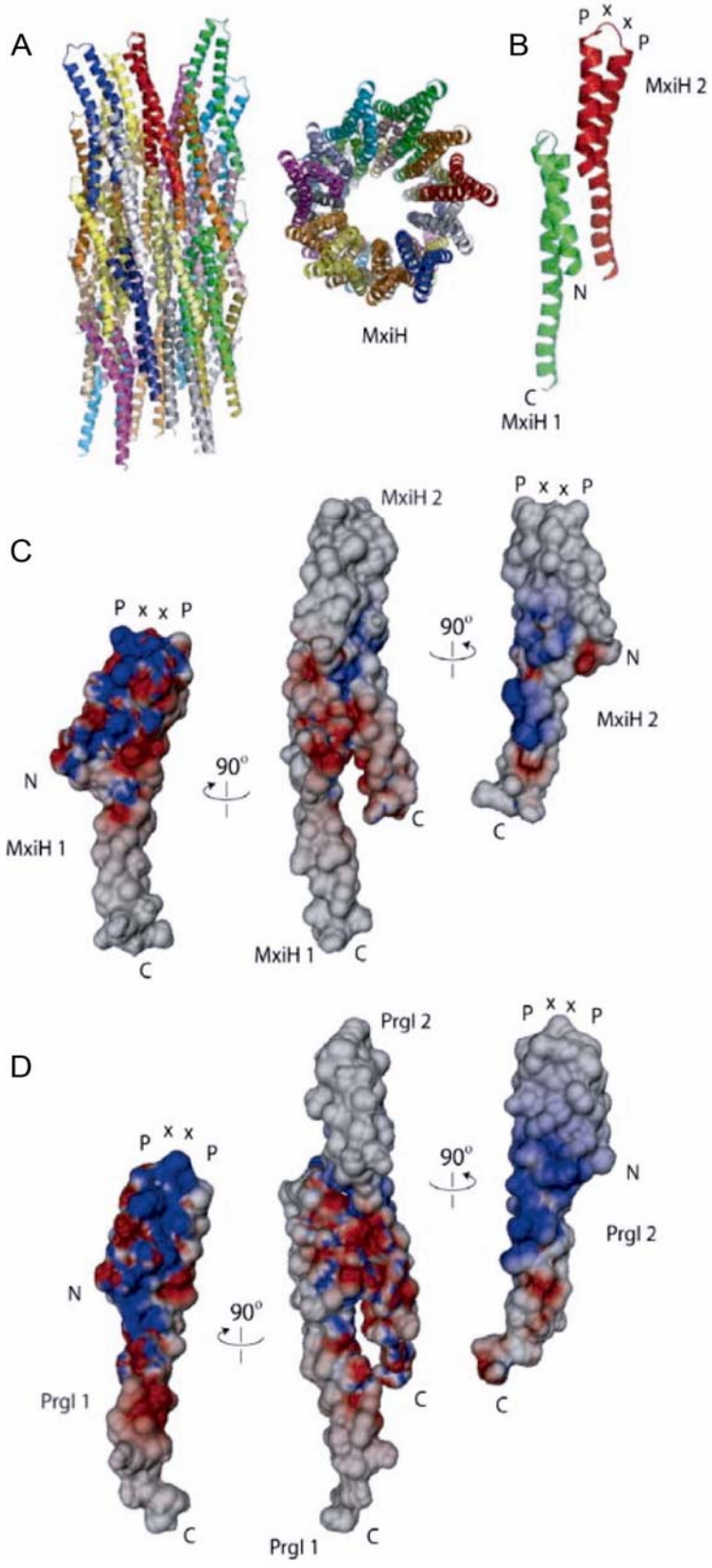
A
MxiH
B
P x x P
MxiH 2
N
C
MxiH 1
C
P x x P
MxiH 2
90°
N
MxiH 2
MxiH 1
C
C
P x x P
N
MxiH 1
C
D
Prgl 2
P x x P
90°
N
Prgl 1
C
P x x P
N
Prgl 1
C
Prgl 2
C

Figure 3-11. (**A**) Model of the *S. flexneri* needle created using the crystal structure of $MxiH^{C\Delta5}$ (PDB code 2CA5) and published needle packing parameters, and visualized with the top towards the host and below towards the bacterium, and along the 25 Å needle channel looking down towards the bacterium. (**B**) Two MxiH molecules within the needle with extensive intermolecular contacts are chosen for electrostatic analysis. (**C**) The two monomers, MxiH1 and MxiH2, are separated from the MxiH dimer and rotated by 90° towards the viewer to reveal the electrostatic potentials of the residues involved in the needle packing interaction. The N-terminal 19 residues of MxiH were removed because they were disordered in the MxiH crystal structure. (**D**) Electrostatic analysis of two PrgI molecules (PrgI1 and PrgI2) in the model of the *S. typhimurium* needle, which was created by superimposing the NMR structure of PrgI into *S. flexneri* needle. The N-terminal 19 residues of PrgI were removed in the calculation of electrostatic potentials because they did not converge into a single family of structures. The Coulombic electrostatic interaction energy is −120 kcal/mol for the MxiH dimer and −23 kcal/mol for the PrgI dimer. Molecular graphics were made using PyMol (A and B) and Dino (DINO: Visualizing Structural Biology (2002). http://www.dino3d.org) (C and D). The N- and the C- termini and the PxxP loop are indicated.[16]

CHARMM. Only the electrostatic potentials of residues involved in the packing interaction are shown in Figure 3-11C and D. The electrostatic maps of MxiH in Figure 3-7C and Figure 3-11C are not identical because the former was calculated from an isolated MxiH structure and the later was that of MxiH in the assembled needle after energy minimization, which changed the side-chain orientations to reduce the steric clash and optimize the MxiH–MxiH contacts in the needle. Thus, the relative side-chain coordinates of MxiH in Figure 3-7C and Figure 3-11C are not identical, and their electrostatic surfaces will not be identical as well. Similarly, the electrostatic maps of the isolated PrgI (Fig. 3-7A) and that of PrgI in the assembled needle after energy minimization (Fig. 3-11D) will not be identical.

For MxiH, there is complementarity between the positively and the negative charged electrostatic surfaces of two MxiH molecules (Fig. 3-11C), whereas, PrgI showed mostly positive-to-positive electrostatic contacts (Fig. 3-11D). In addition, the calculated Coulombic electrostatic interaction energy of the MxiH dimer (Fig. 3-11C) was five times lower than of the PrgI dimer (Fig. 3-11D). Although the packing interactions should be considered in the context of the entire needle, our modeling suggests that the interaction between PrgI molecules would have to be different from that of MxiH to maximize the complementarity between the positive-to-negative electrostatic contacts. Therefore, we conclude that the needle packing interaction in *S. typhimurium* would have to be different from that of *S. flexneri,* even though their needle monomers share 63% primary sequence identity, because the electrostatic surfaces of their needle monomers are radically different from each other.

ACKNOWLEDGEMENT

This work would not be possible without the help of Thenmalarchelvi Rathinavelan and Wonpil Im, and they built the *S. flexneri* needle using the published crystal structure of MixH monomer[13] and needle packing parameters[36; 37]. Thenmalarchelvi Rathinavelan and Wonpil Im also performed the electrostatic analysis of both MxiH and PrgI dimers.

REFERENCES

1. Vugia, D. J., Samuel, M., Farley, M. M., Marcus, R., Shiferaw, B., Shallow, S., Smith, K. & Angulo, F. J. (2004). Invasive *Salmonella* infections in the United States, FoodNet, 1996-1999: incidence, serotype distribution, and outcome. *Clin Infect Dis* **38 Suppl 3**, S149-56.
2. Hensel, M. (2000). *Salmonella* pathogenicity island 2. *Mol Microbiol* **36**, 1015-23.
3. Galan, J. E. & Collmer, A. (1999). Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **284**, 1322-8.
4. Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E. & Holden, D. W. (1995). Simultaneous identification of bacterial virulence genes by negative selection. *Science* **269**, 400-3.
5. Galan, J. E. (2001). *Salmonella* interactions with host cells: type III secretion at work. *Annu Rev Cell Dev Biol* **17**, 53-86.
6. Coburn, B., Sekirov, I. & Finlay, B. B. (2007). Type III secretion systems and disease. *Clin Microbiol Rev* **20**, 535-49.
7. Cirillo, D. M., Valdivia, R. H., Monack, D. M. & Falkow, S. (1998). Macrophage-dependent induction of the *Salmonella* pathogenicity island 2 type III secretion system and its role in intracellular survival. *Mol Microbiol* **30**, 175-88.
8. Hensel, M., Shea, J. E., Waterman, S. R., Mundy, R., Nikolaus, T., Banks, G., Vazquez-Torres, A., Gleeson, C., Fang, F. C. & Holden, D. W. (1998). Genes encoding putative effector proteins of the type III secretion system of *Salmonella* pathogenicity island 2 are required for bacterial virulence and proliferation in macrophages. *Mol Microbiol* **30**, 163-74.

9. Ochman, H., Soncini, F. C., Solomon, F. & Groisman, E. A. (1996). Identification of a pathogenicity island required for *Salmonella* survival in host cells. *Proc Natl Acad Sci U S A* **93**, 7800-4.
10. Takaya, A., Tomoyasu, T., Matsui, H. & Yamamoto, T. (2004). The DnaK/DnaJ chaperone machinery of *Salmonella enterica* serovar Typhimurium is essential for invasion of epithelial cells and survival within macrophages, leading to systemic infection. *Infect Immun* **72**, 1364-73.
11. Kubori, T., Matsushima, Y., Nakamura, D., Uralil, J., Lara-Tejero, M., Sukhan, A., Galan, J. E. & Aizawa, S. I. (1998). Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* **280**, 602-5.
12. Zhang, L., Wang, Y., Picking, W. L., Picking, W. D. & De Guzman, R. N. (2006). Solution structure of monomeric BsaL, the type III secretion needle protein of *Burkholderia pseudomallei*. *J Mol Biol* **359**, 322-30.
13. Deane, J. E., Roversi, P., Cordes, F. S., Johnson, S., Kenjale, R., Daniell, S., Booy, F., Picking, W. D., Picking, W. L., Blocker, A. J. & Lea, S. M. (2006). Molecular model of a type III secretion system needle: Implications for host-cell sensing. *Proc Natl Acad Sci U S A* **103**, 12529-33.
14. Quinaud, M., Ple, S., Job, V., Contreras-Martel, C., Simorre, J. P., Attree, I. & Dessen, A. (2007). Structure of the heterotrimeric complex that regulates type III secretion needle formation. *Proc Natl Acad Sci U S A* **104**, 7803-8.
15. Sun, P., Tropea, J. E., Austin, B. P., Cherry, S. & Waugh, D. S. (2008). Structural characterization of the *Yersinia pestis* type III secretion system needle protein YscF in complex with its heterodimeric chaperone YscE/YscG. *J Mol Biol* **377**, 819-30.
16. Wang, Y., Ouellette, A. N., Egan, C. W., Rathinavelan, T., Im, W. & De Guzman, R. N. (2007). Differences in the electrostatic surfaces of the type III secretion needle proteins PrgI, BsaL, and MxiH. *J Mol Biol* **371**, 1304-14.
17. Wang, Y., Zhang, L., Picking, W. L., Picking, W. D. & Guzman, R. N. D. (2008). Structural dissection of the extracellular moieties of the type III secretion apparatus. *Mol. BioSyst.* **4**, 1176-80.
18. Darboe, N., Kenjale, R., Picking, W. L., Picking, W. D. & Middaugh, C. R. (2006). Physical characterization of MxiH and PrgI, the needle component of the type III secretion apparatus from *Shigella* and *Salmonella*. *Protein Sci* **15**, 543-52.
19. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-93.
20. Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* **278**, 313-52.
21. Grzesiek, S. & Bax, A. (1993). The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115**, 12953-12954.

22. Tolman, J. R., Chung, J. & Prestegard, J. H. (2004). Pure-phase heteronuclear multiple-quantum spectroscopy using field gradient selection *J. Magn. Reson.* **98**, 462-467.
23. Grzesiek, S., Dobeli, H., Gentz, R., Garotta, G., Labhardt, A. M. & Bax, A. (1992). 1H, 13C, and 15N NMR backbone assignments and secondary structure of human interferon-gamma. *Biochemistry* **31**, 8180-90.
24. Muhandiram, D. R. & Kay, L. E. (1994). Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. ser. B* **103**, 203-216.
25. Wittekind, M. & Mueller, L. (1993). HNCACB, a high sensitivity 3D NMR experiment to correlate amide proton and nitrogen resonances with the α-carbon and β-carbon resonances in proteins. *J. Magn. Reson.* **101B**, 201-205.
26. Grzesiek, S. & Bax, A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J Biomol NMR* **3**, 185-204.
27. Wishart, D. S. & Nip, A. M. (1998). Protein chemical shift analysis: a practical guide. *Biochem Cell Biol* **76**, 153-63.
28. Osiecki, J. C., Barker, J., Picking, W. L., Serfis, A. B., Berring, E., Shah, S., Harrington, A. & Picking, W. D. (2001). IpaC from *Shigella* and SipC from *Salmonella* possess similar biochemical properties but are functionally distinct. *Mol Microbiol* **42**, 469-81.
29. Klein, J. R., Fahlen, T. F. & Jones, B. D. (2000). Transcriptional organization and function of invasion genes within *Salmonella enterica* serovar Typhimurium pathogenicity island 1, including the prgH, prgI, prgJ, prgK, orgA, orgB, and orgC genes. *Infect Immun* **68**, 3368-76.
30. Guntert, P. (2004). Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**, 353-78.
31. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham Iii, T. E., Wang, J., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K. & Kollman, P. A. (2002). Amber7. *Amber7*, University of California, San Francisco.
32. Kenjale, R., Wilson, J., Zenk, S. F., Saurya, S., Picking, W. L., Picking, W. D. & Blocker, A. (2005). The needle component of the type III secreton of *Shigella* regulates the activity of the secretion apparatus. *J Biol Chem* **280**, 42929-37.
33. Espina, M., Olive, A. J., Kenjale, R., Moore, D. S., Ausar, S. F., Kaminski, R. W., Oaks, E. V., Middaugh, C. R., Picking, W. D. & Picking, W. L. (2006). IpaD localizes to the tip of the type III secretion system needle of *Shigella flexneri*. *Infect Immun* **74**, 4391-400.
34. Zhang, L., Wang, Y., Olive, A. J., Smith, N. D., Picking, W. D., De Guzman, R. N. & Picking, W. L. (2007). Identification of the MxiH needle protein

residues responsible for anchoring invasion plasmid antigen D to the type III secretion needle tip. *J Biol Chem* **282**, 32144-51.
35. Marlovits, T. C., Kubori, T., Sukhan, A., Thomas, D. R., Galan, J. E. & Unger, V. M. (2004). Structural insights into the assembly of the type III secretion needle complex. *Science* **306**, 1040-2.
36. Cordes, F. S., Komoriya, K., Larquet, E., Yang, S., Egelman, E. H., Blocker, A. & Lea, S. M. (2003). Helical structure of the needle of the type III secretion system of *Shigella flexneri*. *J Biol Chem* **278**, 17103-7.
37. Cordes, F. S., Daniell, S., Kenjale, R., Saurya, S., Picking, W. L., Picking, W. D., Booy, F., Lea, S. M. & Blocker, A. (2005). Helical packing of needles from functionally altered *Shigella* type III secretion systems. *J Mol Biol* **354**, 206-11.
38. Quinaud, M., Chabert, J., Faudry, E., Neumann, E., Lemaire, D., Pastor, A., Elsen, S., Dessen, A. & Attree, I. (2005). The PscE-PscF-PscG complex controls type III secretion needle biogenesis in *Pseudomonas aeruginosa*. *J Biol Chem* **280**, 36293-300.
39. Yonekura, K., Maki-Yonekura, S. & Namba, K. (2003). Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* **424**, 643-50.
40. Samatey, F. A., Matsunami, H., Imada, K., Nagashima, S., Shaikh, T. R., Thomas, D. R., Chen, J. Z., Derosier, D. J., Kitao, A. & Namba, K. (2004). Structure of the bacterial flagellar hook and implication for the molecular universal joint mechanism. *Nature* **431**, 1062-8.
41. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* **4**, 187-217.
42. Im, W., Lee, M. S. & III, C. L. B. (2003). Generalized born model with a simple smoothing function. *J. Comp. Chem.* **21**, 1691-1702.
43. Im, W., Beglov, D. & Roux, B. (1998). Continuum Solvation Model: computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comp. Phys. Comm.* **111**, 59-75.
44. Nina, M., Beglov, D. & Roux, B. (1997). Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. *J. Phys. Chem. ser. B* **101**, 5239-5248.

SUPPLEMENTAL MATERIALS

Table 3-S1. NMR experimental detailes for $PrgI^{C\Delta5}$.

NMR experiment	Bruker Pulse Sequence	Experimental Details	Time*
2D HSQC	hsqcfpf3gpphwg	8 NS* 2048×128 (N) pts	20 min
2D HMQC	blhmqcnoesy2d	8 NS 2048×256(C) pts	1h 30min
3D HNCA	hncagpwg3d	16 NS 2048×54(N)×100(C) pts	1d 4h
3D CBCA(CO)NH	cbcaconhgpwg3d	16 NS 2048×54(N)×100(C) pts	1d 6h
3D HNCO	hncogpwg3d	16 NS 2048×48(N)×50(C) pts	13h
3D HNCACB	hncacbgpwg3d	40 NS 2048×50(N)×100(C) pts	2d 19h
3D HBHA(CO)NH	hbhaconhgpwg3d	16 NS 2048×50(N)×96(H) pts	1d 2h
3D HSQC-NOESY	noesyhsqcf3gp193d	8 NS 2048×44(N)×256(H) pts	1d 9h
3D HMQC-NOESY	blhmqcnoesy3d	8 NS 2048×72(C)×244(H) pts	2d 10h
Heteronuclear ^{1}H-^{15}N NOE	rdhsqcnoef3gpsi	24 NS 2048×360(N) pts	19 h

NS*: number of scans
Time*: total acquisition time

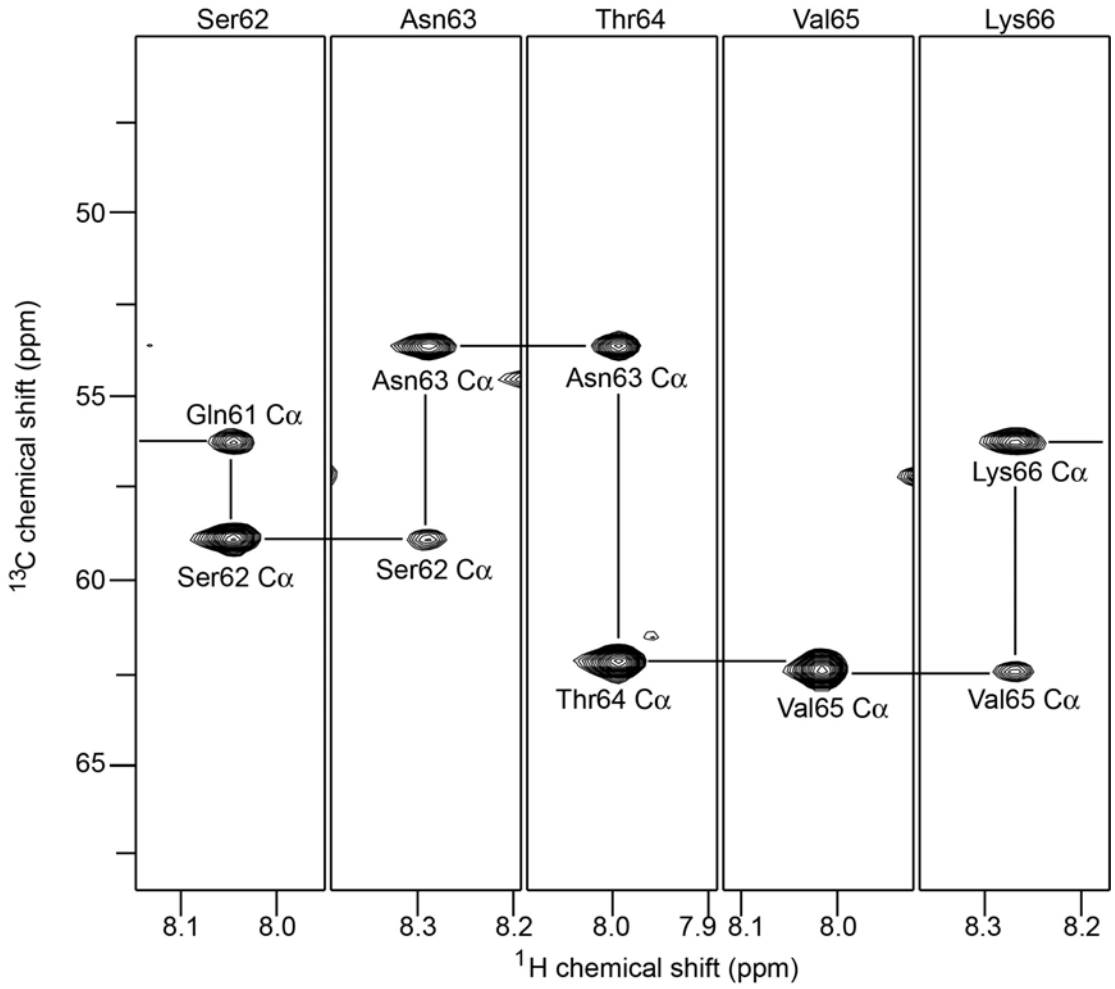

Figure 3-S1. Portion of the 3D HNCA strips of ^{13}C, ^{15}N-labeled PrgI$^{C\Delta5}$, shown as ^{13}C, ^{1}H strips for Ser62-Lys66 (the ^{15}N ppm is on the Z-axis). Each strip shows the C_α of the preceeding and current residue and the amide (^{1}H and ^{15}N) resonances of the current residue, indicating a sequence specific connectivity between these strips. The HNCA data was used together with ^{15}N-HSQC-NOESY, CBCA(CO)NH and HNCACB to complete backbone assignment of PrgI$^{C\Delta5}$.

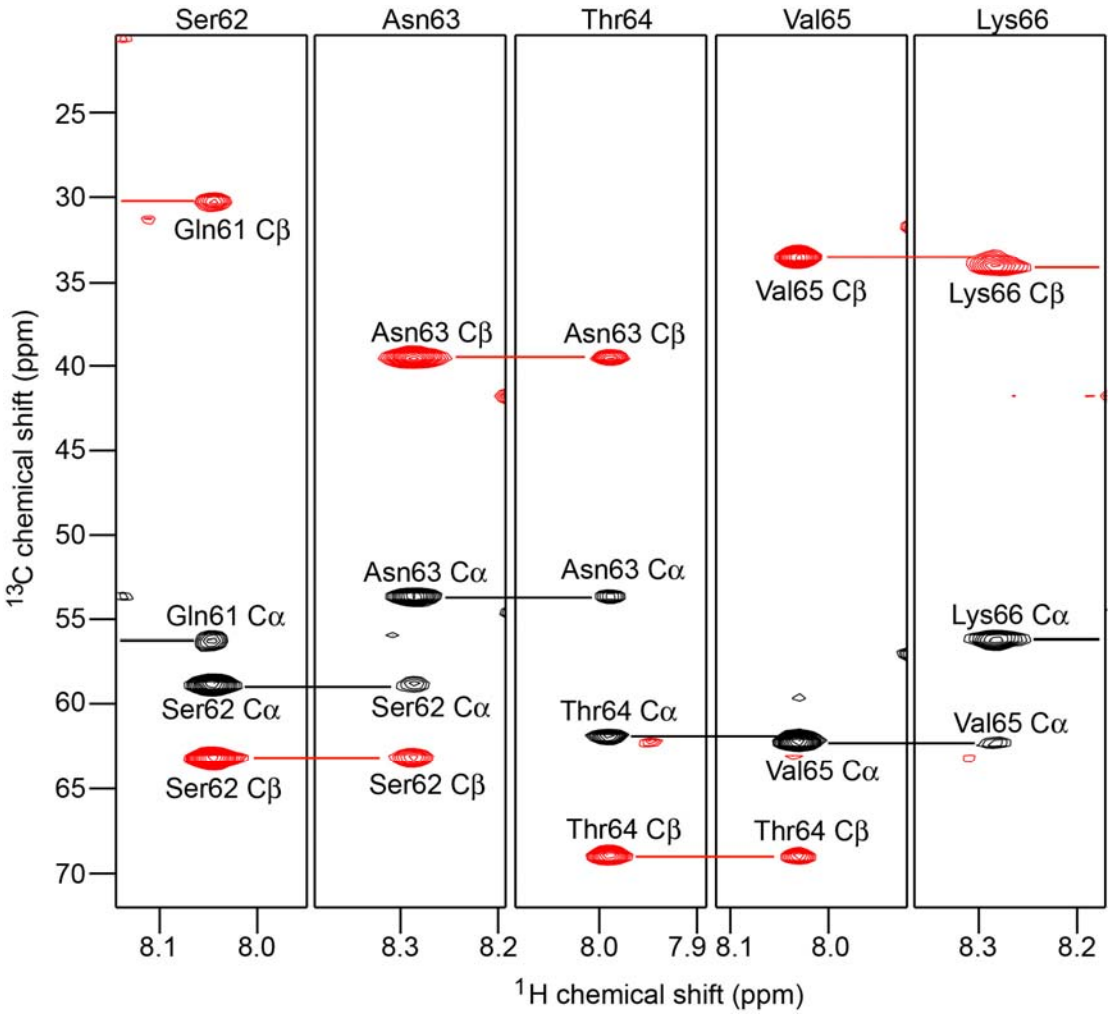

Figure 3-S2. Portion of the 3D HNCACB strips of ^{13}C, ^{15}N-labeled PrgI$^{C\Delta5}$, shown as ^{13}C, ^{1}H strips for Ser62-Lys66 (the ^{15}N ppm is on the Z-axis). Each strip shows the C_α and C_β of the preceeding and current residue and the amide (^{1}H and ^{15}N) resonances of the current residue, indicating sequence specific connectivity between these strips. The C_α and C_β have opposite phases (shown as black and red peaks). In theory, the 3D HNCACB dataset contains all the information needed in assigning all the backbone resonances of a protein. However, this experiment is insensitive, and was used in combination with the more sensitive HNCA, ^{15}N-HSQC-NOESY and CBCA(CO)NH experiments to complete the backbone assignments of PrgI$^{C\Delta5}$.

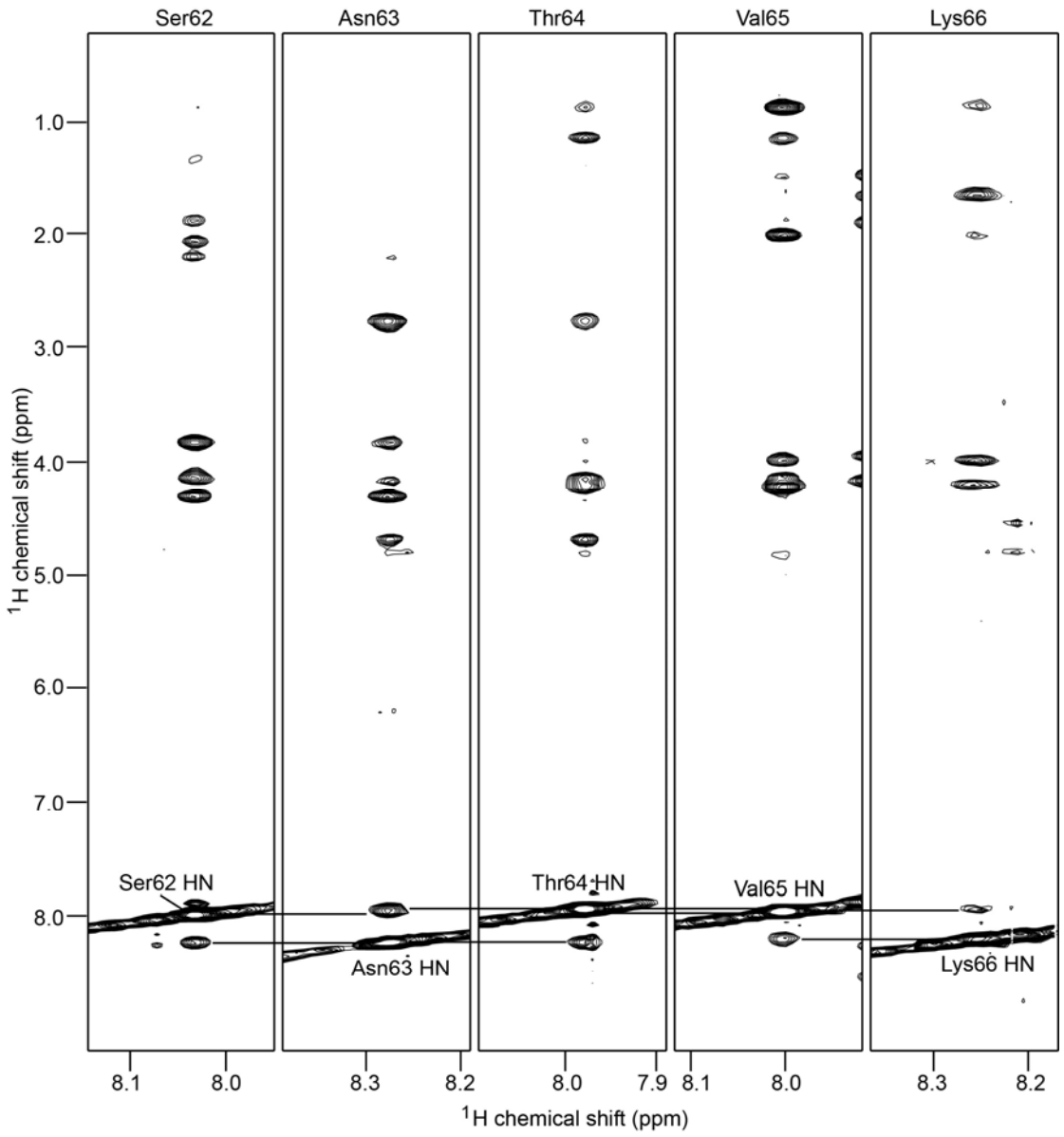

Figure 3-S3. Portion of the 3D ^{15}N-HSQC-NOESY strips of ^{15}N-labeled PrgI$^{C\Delta5}$, shown as ^{1}H, HN strips for Ser62-Lys66 (the ^{15}N ppm is on the Z-axis). Each strip shows all the backbone H_N NOE peaks including its own H_N and the H_Ns of the preceding and following residues, indicating sequence specific connectivity between these strips which can facilitate the backbone assignments.

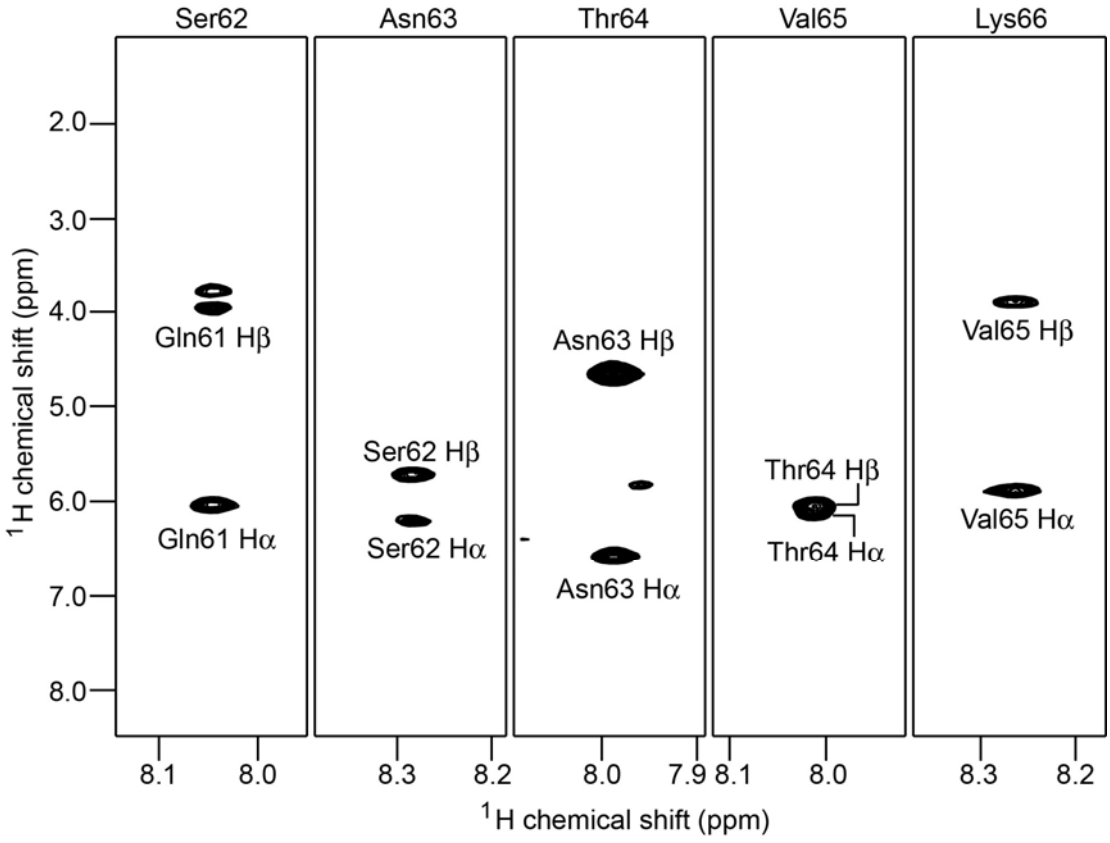

Figure 3-S4. Portion of the 3D HBHA(CO)NH strips of ^{13}C, ^{15}N-labeled PrgI$^{C\Delta5}$, shown as ^{1}H, ^{1}H strips for Ser62-Lys66 (the ^{15}N ppm is on the Z-axis). Each strip shows the H_α and H_β of the preceeding residue and the amide (^{1}H and ^{15}N) resonances of the current residue. The HBHA(CO)NH dataset was collected to assist in the assignment of side chain resonances together with 2D ^{1}H, ^{13}C-HMQC and 3D ^{13}C-HMQC-NOESY of PrgI$^{C\Delta5}$.

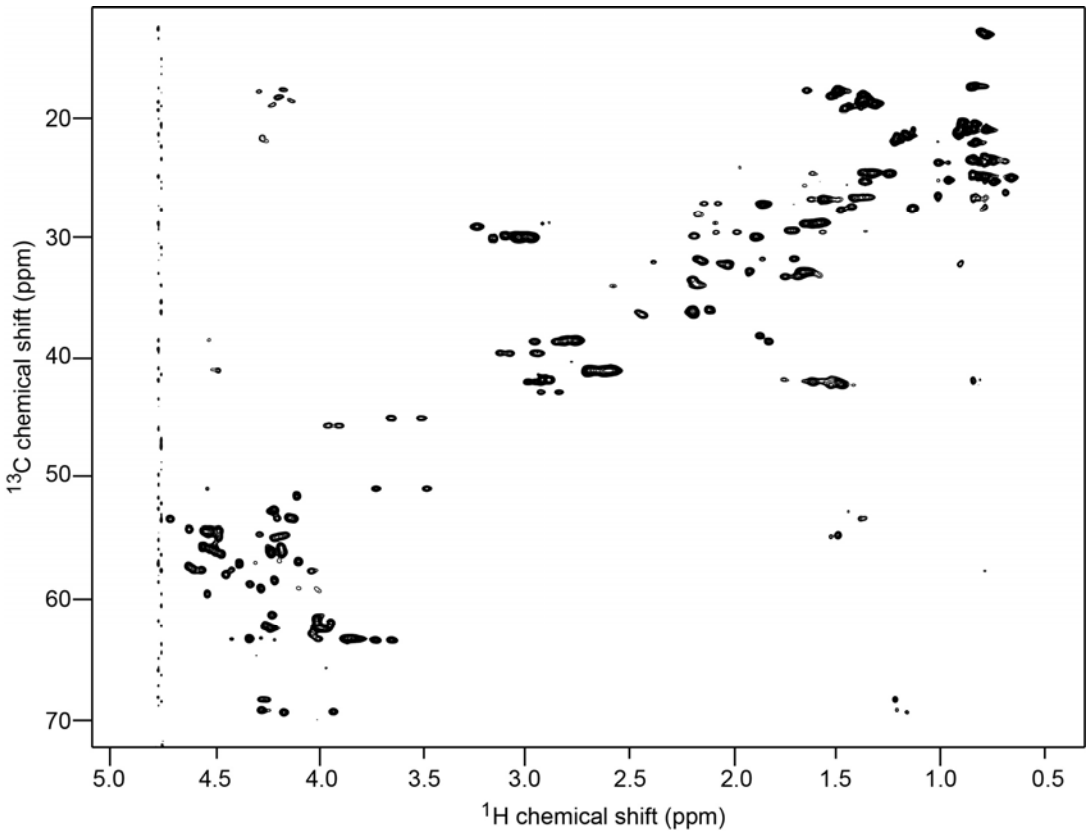

Figure 3-S5A. 2D ^{1}H, ^{13}C-HMQC spectra of ^{13}C, ^{15}N-labeled PrgI$^{C\Delta5}$ in 100% D_2O. The 3D ^{13}C-HMQC-NOESY experiment was based on this 2D HMQC experiment and a third ^{1}H NOE dimension was added. With the help of backbone assignment and 3D HBHA(CO)NH, complete assignment of 2D ^{1}H, ^{13}C-HMQC and 3D ^{13}C-HMQC-NOESY of PrgI$^{C\Delta5}$ were achieved.

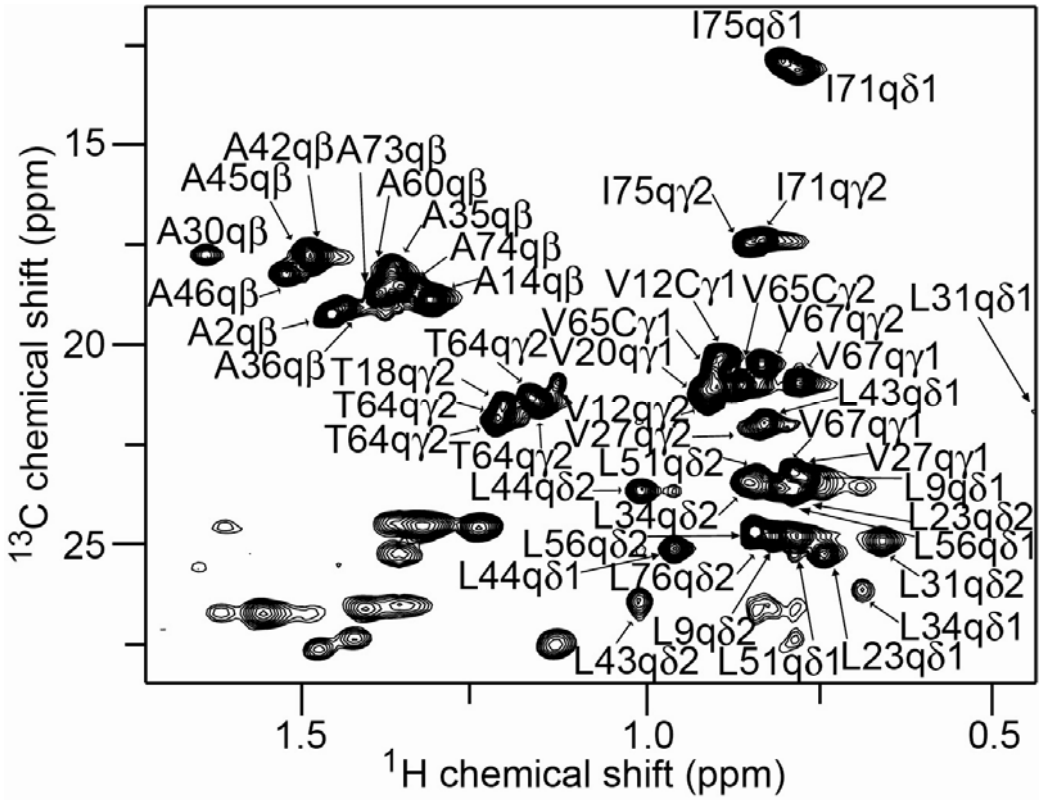

Figure 3-S5B. Methyl group region of 2D ^{1}H, ^{13}C-HMQC spectra of ^{13}C, ^{15}N-labeled $PrgI^{C\Delta5}$ in 100% D_2O.

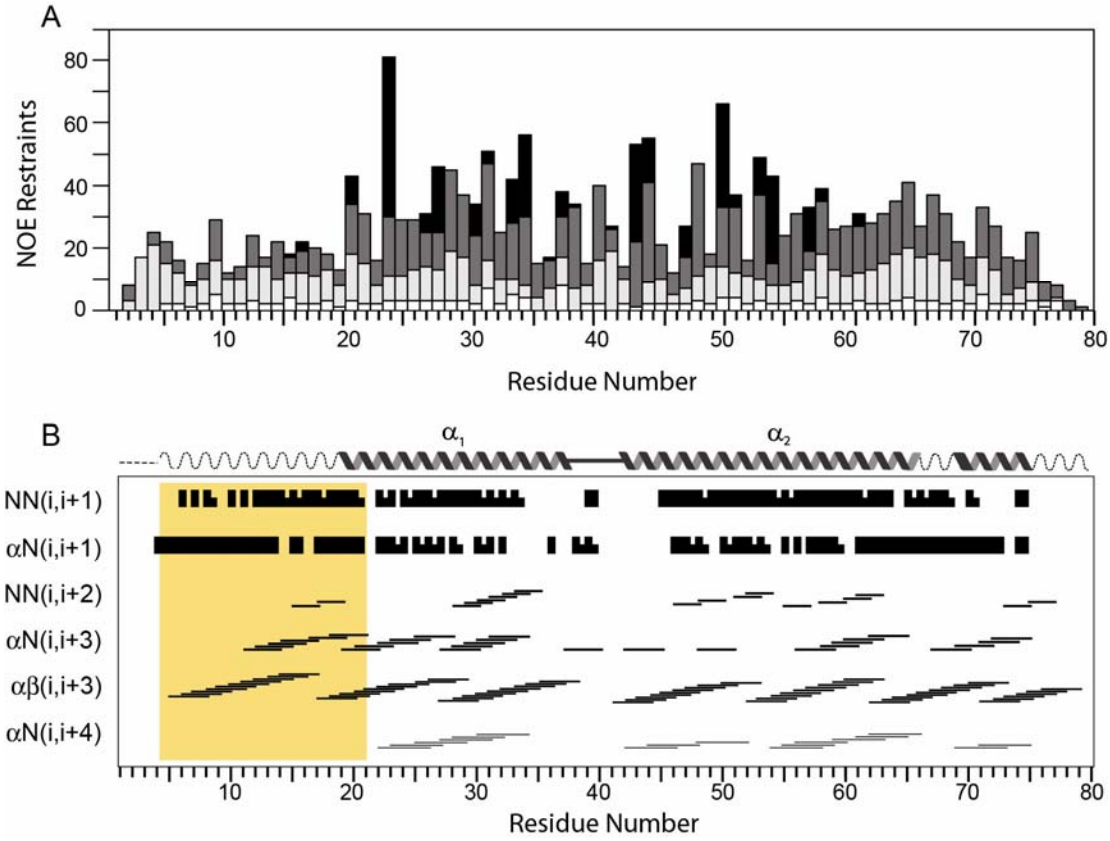

Figure 3-S6. (**A**) The long-range NOE restraints (dark bars) are clustered in the in the α-helical regions of PrgI (V20-V65). The flanking regions (P4-G19 and K66-I75) have medium range (gray bars) and sequential (white bars) NOEs. (**B**) The well defined two-helix bundle in the middle of PrgI (helix α1 and α2) has the characteristic α- helical NOE patterns. The N-terminal region (P4-G19, shaded) also contain medium-range α-helical NOEs.[16]

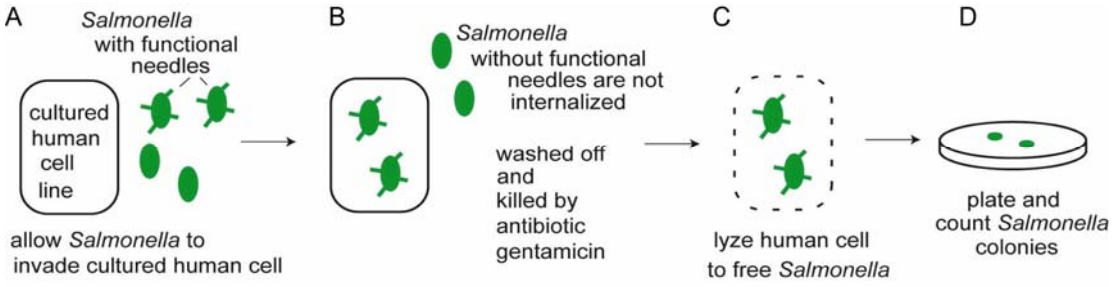

Figure 3-S7. Assay of the ability of *Salmonella* to invade Henle 407, a cultured human epithelial cell line. (**A**) Wild type bacteria can assemble functional needles and therefore can invade and be internalized by Henle 407. (**B**) *Salmonella* with a *prgI* null mutation cannot form functional needles and therefore cannot invade Henle 407, and thus can be washed away and killed with the antibiotic gentamycin. (**C**) Bacteria that had been internalized by Henle 407 are protected from gentamycin, and these bacteria can be freed later by a mild lysis procedure which disrupts the Henle 407 cells but not the bacteria. (**D**) The number of freed bacteria is estimated by serial dilution and plating, and the number of bacterial colonies formed provides a measure of the invasiveness of *S. typhimurium*. A plasmid that expresses PrgI can restore the invasiveness of the *prgI* null mutant and PrgI mutations are introduced in this plasmid to correlate the functional relevance of the NMR structure of PrgI with respect to the ability of *Salmonella* to form functional needles.[16]

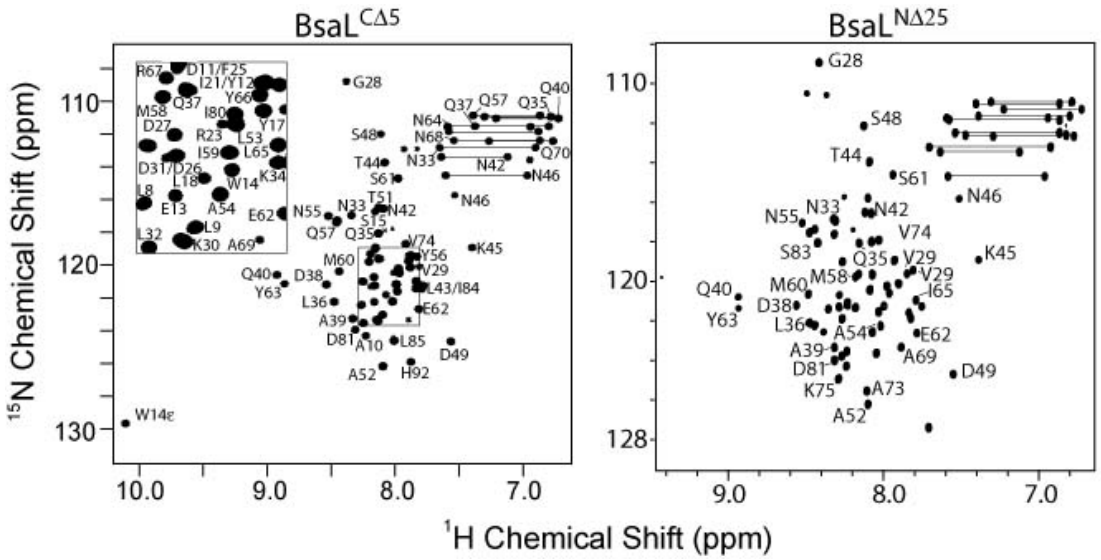

Figure 3-S8. 2D HSQC spectra of two BsaL constructs $BsaL^{C\Delta5}$ and $BsaL^{N\Delta25}$ used for BsaL monomer-monomer interactions. $BsaL^{N\Delta25}$ was soluble and the most of the peaks in the spectra were overlapped with $BsaL^{C\Delta5}$, showing $BsaL^{N\Delta25}$ has the same overall structure as $BsaL^{C\Delta5}$. Many of the specific assignments of $BsaL^{C\Delta5}$ can be transferred to the $BsaL^{N\Delta25}$ because the peaks were in identical positions.

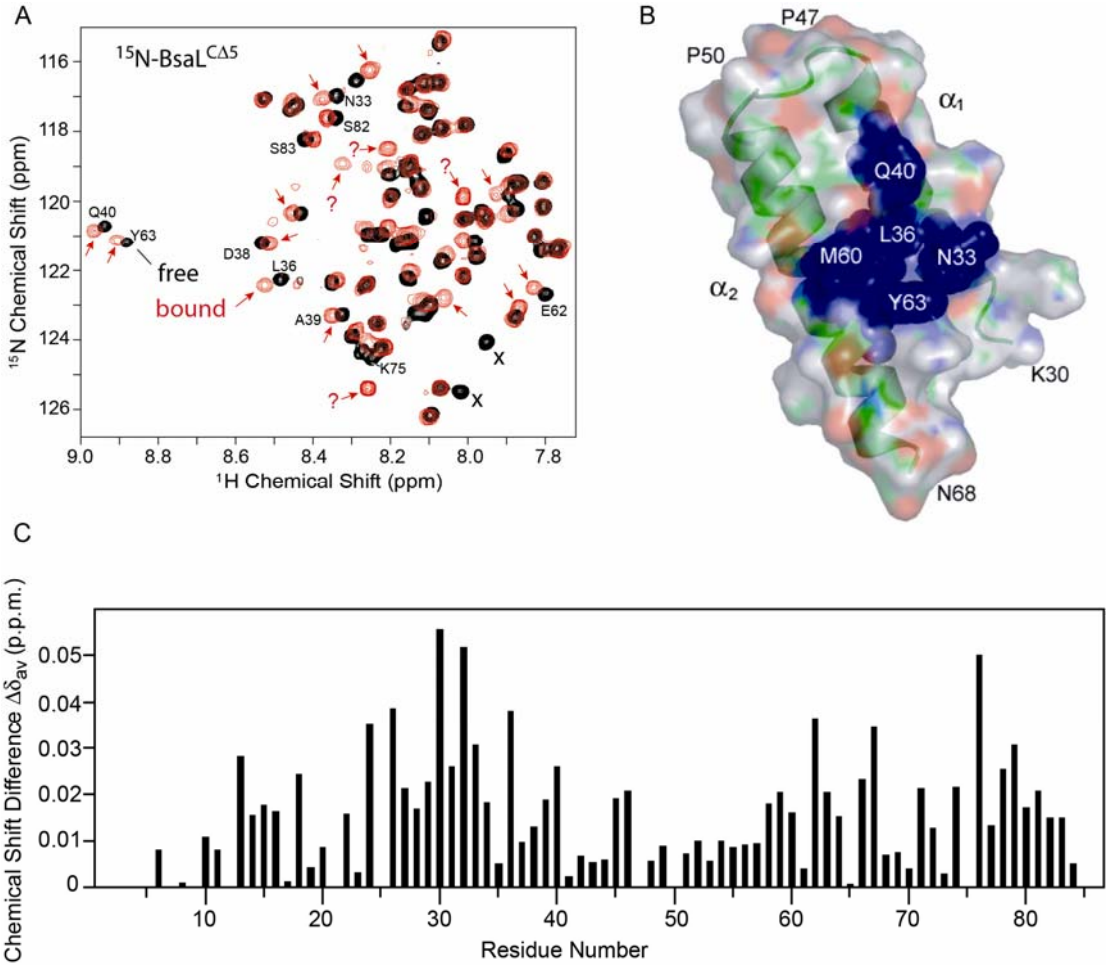

Figure 3-S9. (**A**)Titration of ^{15}N-BsaL^{CΔ5} with increasing amounts of unlabeled BsaL^{NΔ25}. Overlay of two 2D HSQC spectra at two BsaL^{NΔ25}:BsaL^{CΔ5} molar ratios (*black*, 0 and *red*, 8). Most of the residues showed shifts upon dilution, indicating protein-protein interaction. (**B**) Surface representation of residues that showed chemical shift changes upon BsaL-BsaL interaction. Residues Asn33, Leu36, Gln40, Met60 and Tyr63 highlighted in blue clustered at one face of the BsaL structure. (**C**) Chemical shift differences for BsaL^{CΔ5} upon addition of unlabeled BsaL^{NΔ25}. The chemical shift differences were calculated according to the formula $\Delta\delta_{av} = \sqrt{((\Delta\delta_{HN})^2/2 + (\Delta\delta_N/5)^2/2)}$, where $\Delta\delta_{HN}$ and $\Delta\delta_N$ are the amide proton and nitrogen chemical shift differences, respectively. The X-axis is the residue number and the Y axis is the chemical shift differences $\Delta\delta_{av}$ (ppm). Only backbone amides are shown and calculated for chemical shift differences with the missing values being due to incomplete assignments or overlapped peaks.

Chapter 4: Characterization of PrgI Needle Monomer-Monomer Interactions

INTRODUCTION

The atomic structures of needle proteins have provided a wealth of information in explaining biological data related to the phenotypes of needle protein mutants;[1; 2; 3; 4; 5] however, empirical data on how needle monomers interact with each other during needle assembly are still lacking. Furthermore, whether the type III secretion needle is structurally altered during the protein secretion process is still unclear. Efforts on the crystallization of the fully assembled needle remain infeasible due to the non-uniformity in needle length, both for isolated needles and needles reconstituted *in vitro*.[6]

An atomic model of the *S. flexneri* needle was described at 16 Å resolution using the X-ray fiber diffraction and EM-derived needle helical parameters and the crystal structure of the MxiH$^{C\Delta5}$ monomer.[3; 7] However, information on both the MxiH N-terminal 19 residues and C-terminal 5 residues are unclear in the model and it was suggested that the N-terminal residues form a regular α-helix that line inside the needle channel which might have direct contacts with proteins traveling through the needle.[3] It was proposed that the transmission of the invasion signal from the tip to the base might rely on the reorientation of the needle subunits.[3] The protein side

chain orientation, however, is not accurately determined, which makes studies of subtle structural changes occurring in the needle difficult.

In this chapter, we describe the characterization of the needle monomer-monomer interactions by NMR spectroscopy. It will be shown that PrgI$^{C\Delta5}$ is not completely monomeric in solution and there are residual interactions among PrgI$^{C\Delta5}$ monomers. The residues involved in PrgI-PrgI interaction are clustered mainly on the head group including the central PxxP motif, together with a few key residues on both helices (α_1 and α_2). In addition, the heteronuclear ^{1}H-^{15}N NOE data shows that both termini of PrgI$^{C\Delta5}$ are in intermediate rigidity which suggests that partially structured regions might form rigid structures upon protein-protein interaction.

MATERIALS AND METHODS

Protein Expression and Purification

To obtain soluble forms of PrgI, five residues at the carboxyl terminus were deleted to prevent aggregation, giving rise to PrgI$^{C\Delta5}$ as described in chapter 3. Another soluble form of PrgI, GB1-PrgI^{FL}, was constructed by attaching a solubility tag GB1 (B1 domain of S*treptococcal* protein G) followed by a TEV (tobacco etch virus) protease cleavage site, at the amino terminus of PrgI as described (Dalian *et al*. in preparation). Isotopically (^{15}N) labeled protein was overexpressed in *Escherichia coli* BL21(DE3)(DNAY) grown in 1 liter of M9 minimal medium with 1g/L ^{15}N ammonium chloride (Cambridge Isotope Laboratories, Inc.). Bacteria were grown at 37˚C until midlog phase, induced with 1 mM IPTG (isopropyl-β-D-

thiogalactopyranoside) and incubated in 15˚C shaker overnight. The bacterial cells were collected, resuspended in 30 ml of binding buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5 mM imidazole) and lysed by sonication. The bacteria cell lysate was centrifuged at 22,500 g for 15 min, and the supernatant was loaded onto a Ni^{2+} affinity column (Sigma). The recombinant protein was purified by nickel affinity chromatography as described.[4] The His_6-GB1 tag was cleaved by adding 0.08 mM (volume percentile around 4%) TEV protease into purified GB1-$PrgI^{FL}$ and dialyzing the mixture in 8000 molecular mass cut-off dialysis tubing in buffer (50 mM Tris-HCl, pH 8.0, 20 mM NaCl, 0.5 mM EDTA, 1 mM dithiothreitol) for 16 hours at room temperature. The TEV cleaved protein sample was further dialyzed into binding buffer and purified using Ni^{2+} affinity column. All the purified protein samples were concentrated using Amicon ultracel-3k (Millipore) and then dialyzed into NMR buffer (10 mM sodium phosphate, pH 6.5, 10 mM NaCl) before use and the protein concentration was estimated using absorbance at 280 nm.

NMR Spectroscopy

Uniformly ^{15}N-labeled $PrgI^{C\Delta5}$ and unlabeled GB1-$PrgI^{FL}$ were dissolved in NMR buffer as described above. NMR data were acquired at 25°C using a Bruker Avance 800 MHz spectrometer equipped with a cryogenic Z-axis pulse field gradient probe. The sequential backbone assignments of ^{13}C, ^{15}N-labeled GB1-$PrgI^{FL}$ (Fig. 4-S1A and S1B) and ^{13}C, ^{15}N-labeled GB1-$BsaL^{FL}$ (Fig. 4-S2A and S2B) were obtained using the following triple-resonance experiments: HNCA[8] and CBCA(CO)NH[8]. In

the heteronuclear ^{1}H-^{15}N NOEs experiment, relaxation recycle delays of 5 seconds before the ^{1}H saturations of 3 seconds were applied (Table 2-S1 and 3-S1).[9; 10; 11] Chemical shift mapping experiment was done by acquiring a series of 2D ^{1}H-^{15}N HSQC spectra for each titration point. NMR data were processed using NMRPipe[12] and analyzed using NMRView[13]. The concentration of purified ^{15}N-PrgI$^{C\Delta5}$ and unlabeled GB1-PrgI^{FL} were estimated by UV absorbance at 280nm to be 0.97 mM and 0.54 mM, respectively. In order to exclude the interference of concentration effect, the chemical shift mapping experiments were done using fixed amount of ^{15}N-PrgI$^{C\Delta5}$ and mixing with various amount of unlabeled protein GB1-PrgI^{FL} in different NMR tubes and by compensating the volume using NMR buffer at different GB1-PrgI^{FL}:PrgI$^{C\Delta5}$ volume ratios of 1:0, 1:1, 1:3, 1:6 and 1:11. For the ^{15}N-PrgI$^{C\Delta5}$ dilution experiment, the protein was simply diluted at different dilution factors of 0, 3, 5 and 12. In order to test the function of PrgI C-terminal 5 residues, another titration was done using ^{15}N-GB1-PrgI^{FL} upon addition of unlabeled PrgI$^{C\Delta5}$ at PrgI$^{C\Delta5}$:GB1-PrgI^{FL} molar ratios of 0 and 8.1. In order to exclude the concentration effect of GB1-PrgI^{FL}, the concentration of ^{15}N-GB1-PrgI^{FL} was kept the same in all smples used in the titration. The dilution of the C-terminal deletion mutant ^{15}N-PrgI$^{C\Delta10}$ into NMR buffer was done at dilution factors of 0 and 10. Prior to titration, all the purified protein samples were dialyzed into the same buffer to assure that the chemical shift changes were only due to protein-protein interactions and not to changes in solution condition upon mixing.

Mutagenesis of *prgI*

The *prgI* site specific mutations were introduced by PCR using the Stratagene Quickchange kit in two plasmids: *(i)* pET-22B-*prgI*$^{C\Delta5}$, used to overexpress recombinant His-tagged PrgI$^{C\Delta5}$ in *E. coli* BL21(DE3)(DNAY) cell, and *(ii)* pRK2-*prgI*, used to express full length PrgI for *Salmonella* invasion assays (see below). All mutations were confirmed by DNA sequencing.

Paramagnetic Spin Labeling Experiment

The cysteine-specific spin labeling probe 4-Maleimido-TEMPO (Sigma) was attached to the PrgI$^{C\Delta5}$ N-terminal cysteine mutant N3C. There is no cysteine residue in PrgI$^{C\Delta5}$, so the N3C cysteine mutant was introduced using the Statagene Quickchange kit. The purified PrgI$^{C\Delta5}$N3C formed dimers in solution and 10 mM DTT was added and incubated overnight to reduce the disulfide bond. The protein was dialyzed into NMR buffer (10 mM sodium phosphate pH 6.5 and 10 mM NaCl) before spin labeling.[14] The preparation of the spin-labeled protein was as follows: ~0.4 mM ^{15}N-PrgI$^{C\Delta5}$-N3C was incubated with 10-fold excess of 4-Maleimido-TEMPO at room temperature for 30 min.[14] The reaction was quenched by the addition of 10 mM β-mercaptoethanol, and the spin-labeled protein was purified using a NAP-5 column Sephadex G-25 (GE Healthcare) to remove excess spin probe. NMR data was collected using a regular ^{1}H, 15N-HSQC pulse sequence for the spin-labeled and unlabelled proteins. The peak intensities were compared between spin labeled protein and the unlabeled protein, and plotted by the residue number.

RESULTS

Concentration Effect

After completion of the NMR structure of *Salmonella* PrgI$^{C\Delta5}$ monomer[4], the question addressed is how needle monomers interact with each other during needle assembly. NMR chemical shift mapping was used to identify PrgI residues that are involved in needle monomer-monomer interactions. A soluble form of full-length PrgI was made by fusion with a solubility tag, the GB1 domain, to the N-terminus of the needle protein PrgI. The GB1 tagged full-length PrgI, GB1-PrgI^{FL}, is highly soluble in solution.

Chemical shift mapping used in characterizing protein-protein interactions involves titration of an unlabeled protein into a ^{15}N-labeled protein partner. Peak shifts in the 2D ^{1}H, ^{15}N-HSQC spectra upon titration indicates possible interaction. Surprisingly, when ^{15}N-labeled PrgI$^{C\Delta5}$ were titrated with unlabeled proteins or with NMR buffer, similar peak shifts were observed (Fig. 4-1A), indicating these peak shifts resulted from the perturbation of the protein-protein interaction among PrgI$^{C\Delta5}$ monomers. This residual interaction among soluble PrgI$^{C\Delta5}$ monomers could be attenuated by increasing ionic strength, for example, by increasing salt concentration from 0 mM to 150 mM. Consistent with the results of chemical shift mapping, paramagnetic spin labelling data showed that the N-terminal spin probe also affected not only residues close to the N-terminus of PrgI but also residues on the helix α_1 far away from the N-terminus (Fig. 4-2), suggesting the N-terminus might be involved in

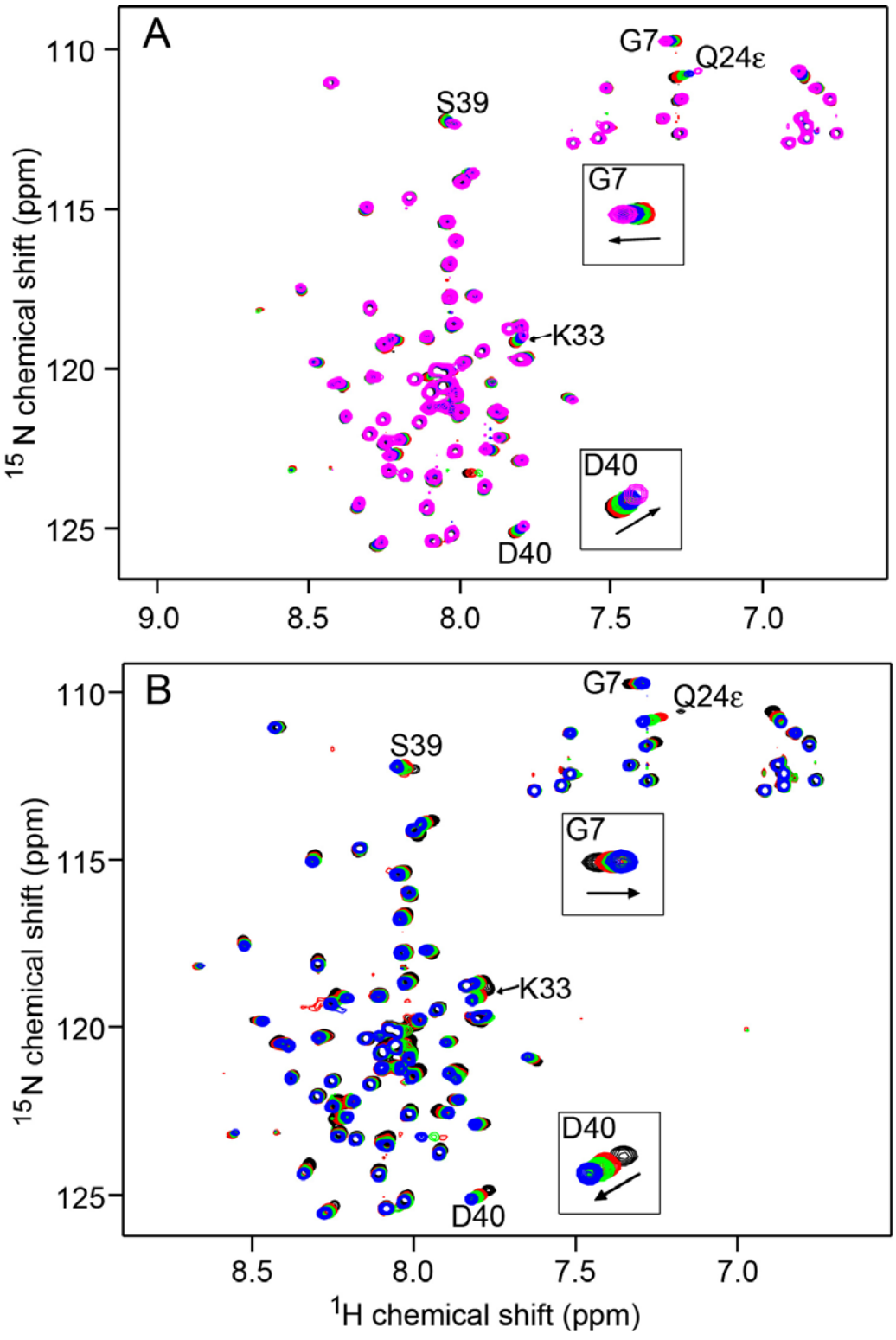
A
G7
Q24ε
S39
G7
K33
D40
D40
110
115
120
125
9.0
8.5
8.0
7.5
7.0
^{15}N chemical shift (ppm)
B
G7
Q24ε
S39
G7
K33
D40
D40
110
115
120
125
8.5
8.0
7.5
7.0
^{15}N chemical shift (ppm)
^{1}H chemical shift (ppm)

Figure 4-1A. (**A**) Dilution of ^{15}N-PrgI$^{C\Delta5}$ with increasing amounts of NMR buffer. Overlay of four 2D HSQC spectra at various dilution factors (*black*, 0; *red*, 3.0; *green*, 5.0; *blue*, 12.0). Most of the residues showed shifts upon dilution. (**B**) Titration of ^{15}N- PrgI$^{C\Delta5}$ with increasing amounts of unlabeled GB1-PrgI^{FL}. Overlay of five 2D HSQC spectra at various GB1-PrgI^{FL}:PrgI$^{C\Delta5}$ molar ratios (*black*, 0; *red*, 0.56; *green*, 1.67; *blue*, 3.34; *magenta*, 6.12). Clear shifts were observed upon addition of unlabeled GB1-PrgI^{FL} and the peaks shifted in the totally opposite direction with the shifts observed in dilution experiment. Peaks of individual residues G7 and D40 showing shifts are zoomed in and pointed out for clarity.

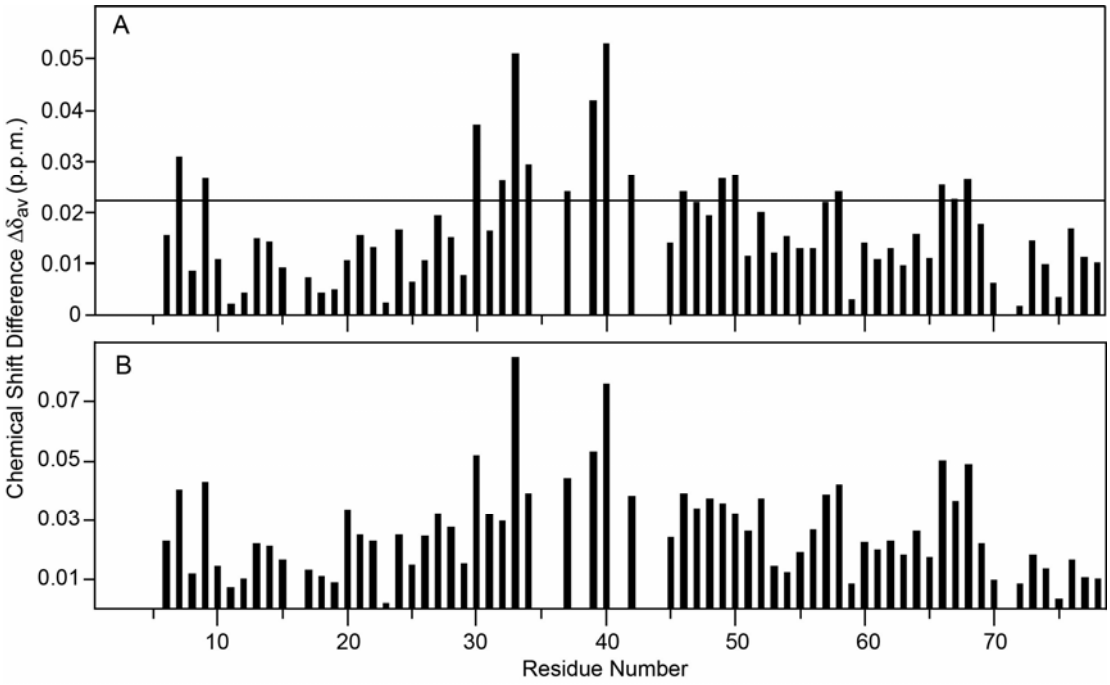

Figure 4-1B. Chemical shift differences for (**A**) PrgI$^{C\Delta5}$ upon addition of unlabeled GB1-PrgI^{FL} and (**B**) titration of buffer comparing two different titration points with BsaL$^{C\Delta5}$:GB1-PrgI^{FL} molar ratios of 0 and 6.12 and dilution factor of 0 and 10. The chemical shift differences were calculated according to the formula $\Delta\delta_{av} = \sqrt{((\Delta\delta_{HN})^2/2 + (\Delta\delta_N/5)^2/2)}$, where $\Delta\delta_{HN}$ and $\Delta\delta_N$ are the amide proton and nitrogen chemical shift differences, respectively. The X-axis is the residue number and the Y axis is the chemical shift differences $\Delta\delta_{av}$ (ppm). Only backbone amides are shown and calculated for chemical shift differences with the missing values being due to incomplete assignments or overlapped peaks.

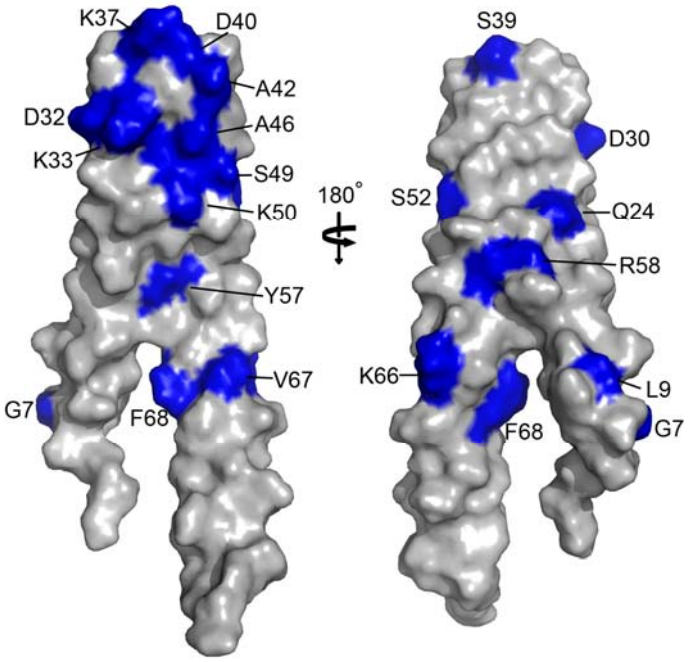

Figure 4-1C. Surface representation of the residues of PrgI$^{C\Delta5}$ (shown in blue) that are perturbed upon PrgI-PrgI interaction. The residues don't cluster into one site and the PrgI-PrgI interaction is involved almost all the parts of PrgI, possibly C-terminal 5 residues as well.

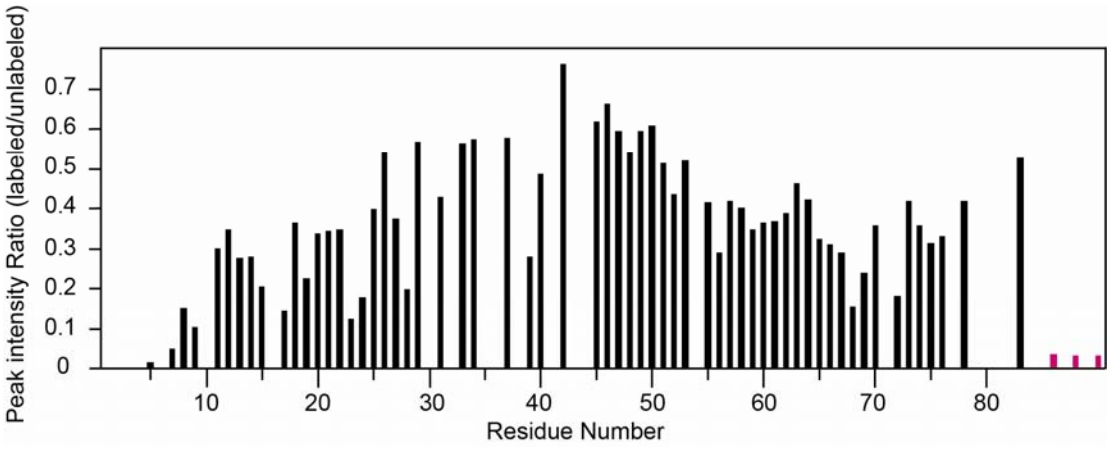

Figure 4-2. The peak intensity ratio of spin labeled versus unlabeled ^{15}N-PrgI$^{C\Delta5}$-N3C mutant. Due to the paramagnetic effect, peak intensity of residues close to the N-terminus was affected by the spin label. In addition to that, a stretch of residues close to C-terminus (F68-D72) were also affected. Side chains of some residues Gln24, Gln26 and Asn55 (shown in red) were greatly affected by the spin probe.

the PrgI-PrgI interaction. The spin labeling data also showed that there is a dip around residue Phe68, indicating C-terminal tail, especially the region around Phe68, is close in distance with the N-terminus (Fig. 4-2). In order to observe real peak shifts for itrations between $PrgI^{C\Delta5}$ and other proteins, the residual interaction among $PrgI^{C\Delta5}$ molecules (concentration effect) must be excluded. This concentration effect was also observed for $MxiH^{C\Delta5}$ (Fig. 4-S3) but not $BsaL^{C\Delta5}$ (Fig. 4-S4).

$PrgI^{C\Delta5}$ Residues Perturbed in PrgI-PrgI Titrations

In order to exclude the concentration effect, we carefully kept the same concentration of ^{15}N-labeled $PrgI^{C\Delta5}$ and varied the amount of unlabeled GB1-$PrgI^{FL}$ in each titration point. The 2D ^{1}H-^{15}N-HSQC spectra show that a subset of the backbone and side-chain amide resonances progressively shift upon addition of increasing amount of unlabeled GB1-$PrgI^{FL}$ (Fig. 4-1A). Small shifts of those perturbed residues imply that the interaction between PrgI monomers is in fast exchange NMR time scale, and weak interaction in the order of 0.7 mM (Fig. 4-3). The $PrgI^{C\Delta5}$ dilution with NMR buffer also showed a weak PrgI-PrgI interaction, however, in a relative higher affinity in the order of 0.07 mM. Additionally, it was noticed that the peaks shifted in the opposite direction relative to the data of $PrgI^{C\Delta5}$ titrated with GB1-$PrgI^{FL}$ compared with the $PrgI^{C\Delta5}$ dilution data (Fig. 4-1B). Considering the opposite titration effect on the total protein concentration of different forms of soluble PrgI, results of these two datasets are consistent. Similar results were shown by analysis of the chemical shift differences of these two datasets (Fig 4-1B).

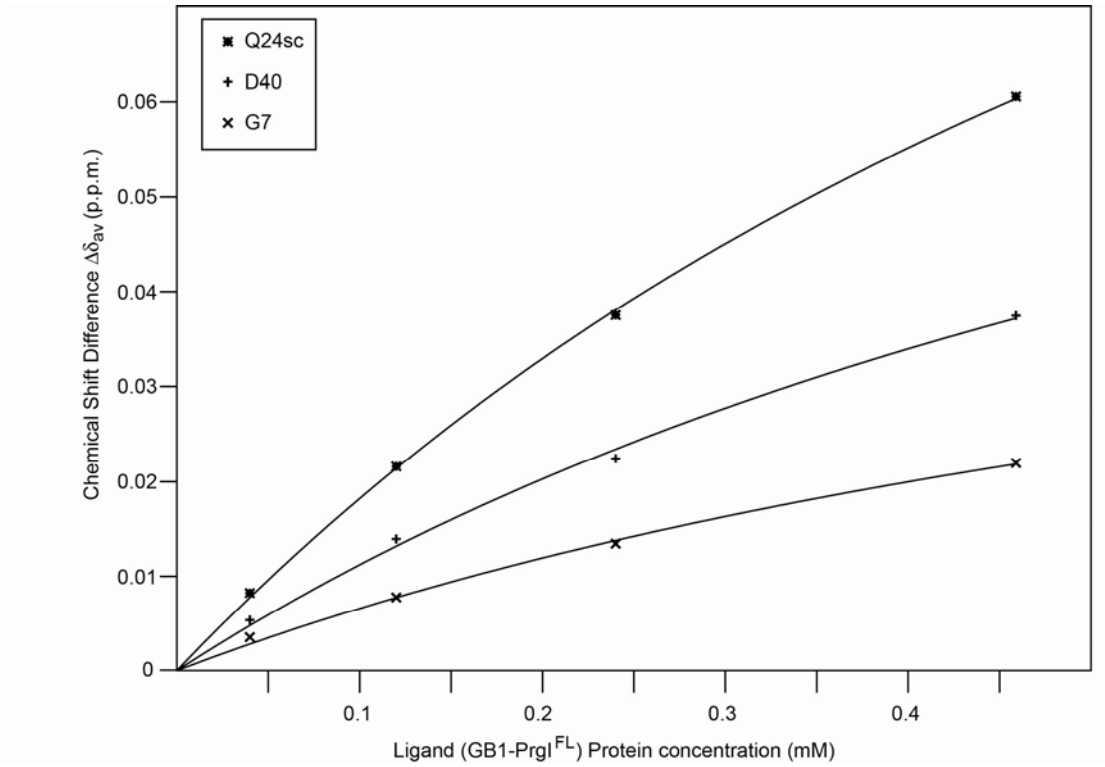

Figure 4-3. Simultaneous fit performed on chemical shift perturbations of 2 residues (Gly7 and Asp40) backbone and 1 residue (Gln24) side-chain HN cross-peaks in the $PrgI^{C\Delta5}$:GB1-$PrgI^{FL}$ titration according to the various concentrations of unlabeled GB1-$PrgI^{FL}$ using following equation.[15]

$$\Delta\text{dppm} = B_{\max}\frac{[\text{Pt}] + [\text{Lt}] + K_D - \sqrt{([\text{Pt}] + [\text{Lt}] + K_D)^2 - 4[\text{Pt}][\text{Lt}]}}{2[\text{Pt}]}$$

where Pt, Lt, K_D and B_{max} are the concentration of protein titrated into, the protein concentration of prtein titrated with, dissociation constant, and the chemical shift deviation at saturation, respectively.

We found that the PxxP motif residues Ser39 and Asp40 which showed significant chemical shift changes are most likely involved in PrgI-PrgI interaction. Residues close to the PxxP motif (Ala30, Asp32, Lys33, Lys37, Ala42, Ala46, Ser49 and Lys50) also showed significant chemical shift changes (Fig. 4-1B). These residues clustered on a surface of the PrgI head group (Fig. 4-1C). In the second helix of PrgI, a series of residues including Ser52, Tyr57, Arg58, Lys66, Val67 and Phe68 also showed relatively significant chemical shift changes, indicating that these residues on the second helix are involved in the PrgI monomer-monomer interaction. The side chain of residue Gln24 showed the most significant chemical shift change, but Gln24 is not included in any group of residues mentioned above. It was surprising that N-terminal residues Gly7 and Leu9 also showed significant changes in the titration. These results suggest more than one binding site is involved in the PrgI monomer-monomer interaction.

A C-terminal deletion of 10 residues was made, $PrgI^{C\Delta10}$, and examined to characterize the residual interaction (Fig. 4-4A). The dilution of ^{15}N-$PrgI^{C\Delta10}$ with NMR buffer showed head group residues (Ala30, Lys33, Ser39 and Asp40) close to the PxxP motif still showed concentration effect and these residues are involved in the residual $PrgI^{C\Delta10}$ monomer interactions (Fig. 4-4B). However, residues Gly7 and Leu9, close to the N-terminus, and residues Lys66, Phe68, close to the C-terminus, coincidently showed small shift change differences compared to the $PrgI^{C\Delta5}$ dilution, suggesting possible interactions among those two stretches of residues (Fig. 4-4B). This is consistent with the spin labeling data described above (Fig. 4-2).

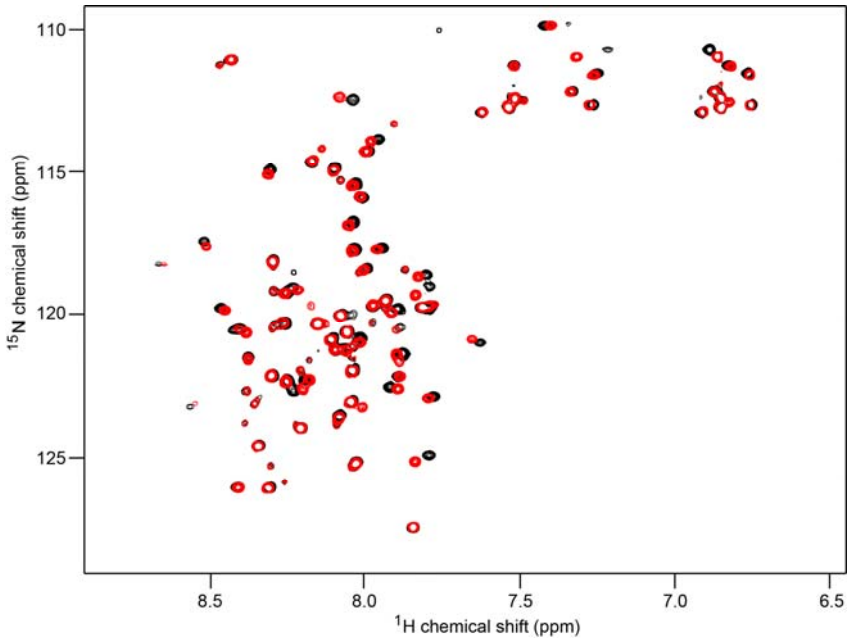

Figure 4-4A. Dilution of ^{15}N-PrgI$^{C\Delta10}$ with increasing amounts of NMR buffer. Overlay of two 2D HSQC spectra at various dilution factors (*black*, 0 and *red*, 10.0). Most of the residues showed peak shifts upon dilution.

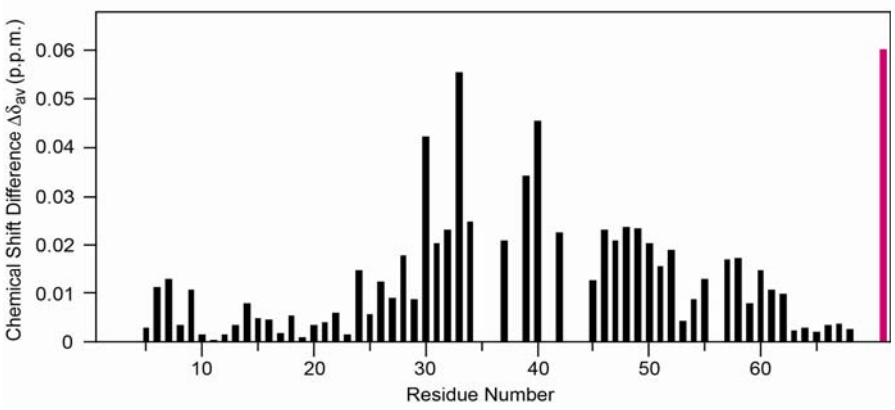

Figure 4-4B. Chemical shift differences for PrgI$^{C\Delta10}$ upon dilution with buffer at dilution factor of 0 and 10. The chemical shift differences were calculated as described in the Material and Methods section and plotted on the y-axis. Backbone amides (shown in black) are shown, and the side chain of residue Gln24 (shown in red) showed significant chemical shift difference. Missing values are due to incomplete assignments or overlapped peaks.

Heteronuclear ^{1}H-^{15}N NOE of PrgI$^{C\Delta5}$ and BsaL$^{C\Delta5}$

Previous NOE data suggest both the N- and C- termini of PrgI adopt weak α-helical character, which is consistent with the low convergence of the NMR structures for the N- and C-termini.[4] Steady-state heteronuclear ^{1}H-^{15}N NOE was used to quantify the degree of internal protein dynamics by the analysis of the fast motions of protein backbone dynamics in nano-second to pico-second range. The core domain two helices showed heteronuclear ^{1}H-^{15}N NOE values (>0.7) which indicates rigid structure. The relatively low heteronuclear ^{1}H-^{15}N NOE values (0.4< NOE <0.6) of both the termini imply they are in medium range mobility (Fig. 4-5A). Here, the N-terminus boundary defines residues from Tyr8 to Asn22 and C-terminus boundary defines residues from Gln61 to the end of the protein. Similarly, both termini of BsaL$^{C\Delta5}$ also adopt medium range rigidity (Fig. 4-5B).

DISCUSSION

PrgI is a small acidic protein that can self-associate into an ordered needle structure on the *Salmonella* surface. It was previously determined that deletion of C-terminal 5 residues prevented the aggregation of PrgI (PrgI$^{C\Delta5}$ is soluble and monomeric in solution).[4] However, peak shifts were observed in the PrgI$^{C\Delta5}$ dilution experiment, suggesting there is residual interaction between PrgI$^{C\Delta5}$ monomers. The previously reported *Shigella* needle model showed there are 7 identical MxiH monomers surrounding a single MxiH monomer and multiple contacts are involved in

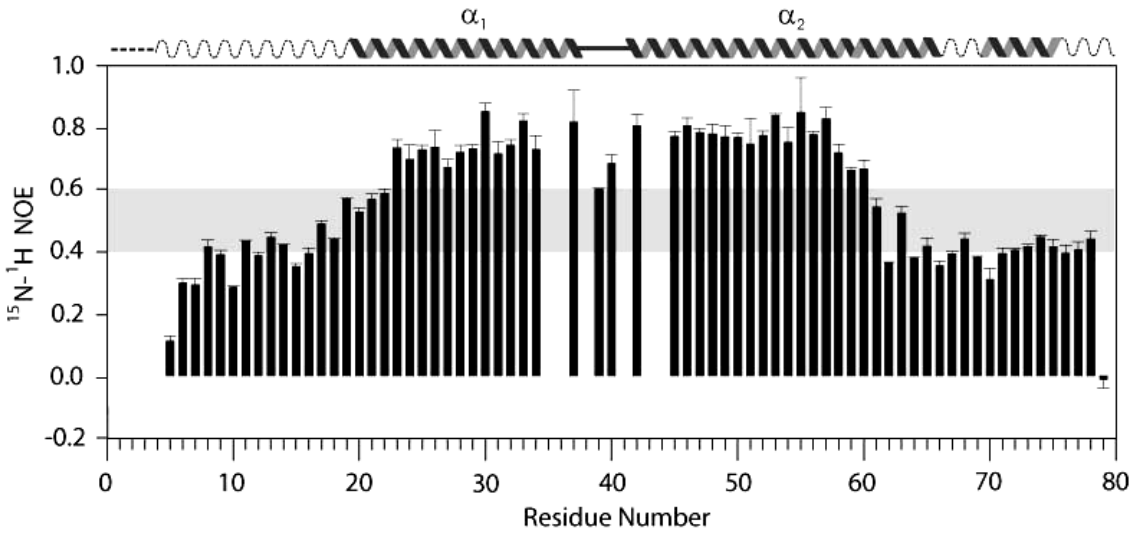

Figure 4-5A. Heteronuclear ^{1}H-^{15}N nuclear Overhauser enhancement of PrgI$^{C\Delta5}$. Small or negative values indicate a high degree of flexibility. The horizontal bar at 0.4-0.6 marks an approximate boundary between flexible (below) and rigid (above) amides on a subnanosecond time scale.

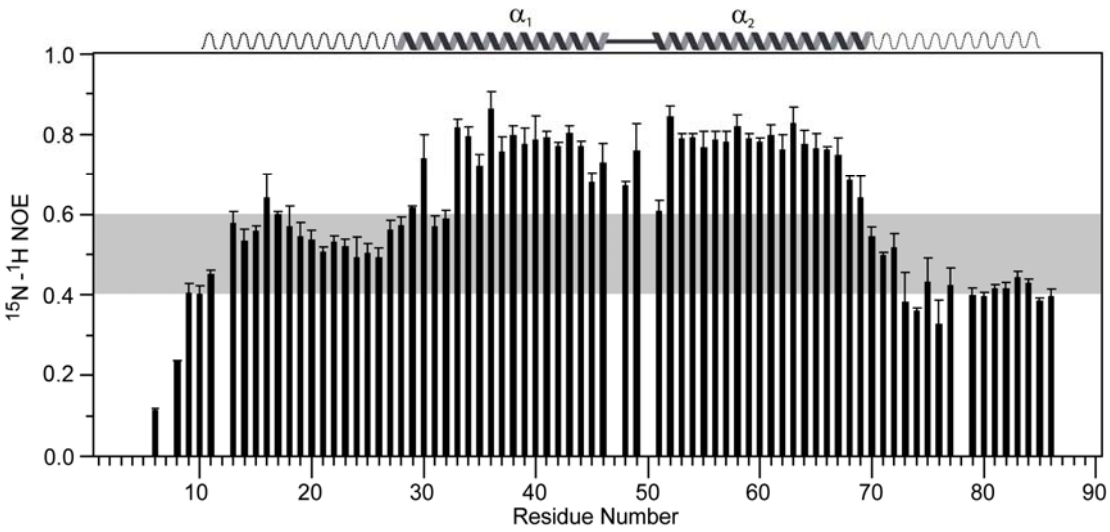

Figure 4-5B. Heteronuclear ^{1}H-^{15}N nuclear Overhauser enhancement of BsaL$^{C\Delta5}$. Small values indicate a high degree of flexibility. The horizontal bar at 0.4-0.6 marks an approximate boundary between flexible (below) and rigid (above) amides on a subnanosecond time scale.

the needle monomer-monomer interaction.[3] The NMR chemical shift mapping data of titrating PrgI$^{C\Delta5}$ with GB1-PrgI^{FL} suggests that the PrgI head group residues, including the PxxP motif, along with some residues on both helices, are involved in the needle monomer-monomer interaction. These residues are not clustered in more than one binding sites, indicating that multiple sites are likely to be involved in this protein-protein interaction. Many of the perturbed residues are charged residues, suggesting electrostatic interaction could be important in PrgI-PrgI interaction. Additional hydrophobic effect might contribute to this interaction as well because a few hydrophobic residues were also perturbed in the titration data. Our mapping data also show that PrgI monomer-monomer interaction *in vitro* is weak. In the assembled needle, the high local concentration of needle monomers probably compensates for the weak interaction observed *in vitro* and it becomes strong enough to hold the needle monomers together. In order to achieve the fine control of needle assembly, strong interactions between needle monomers is probably not necessary. The weak interactions between PrgI monomers are corroborated by the low dissociation constants fitted using the data of PrgI$^{C\Delta5}$ titrated with GB1-PrgI^{FL} and PrgI$^{C\Delta5}$ diluted with NMR buffer. Nevertheless, our data identified multiple binding sites involved in the weak PrgI-PrgI interaction, suggesting *Salmonella* needle is assembled by the binding of a single PrgI monomer with a number of other PrgI monomers, similar to the proposed *Shigella* needle model.[3]

The relatively low heteronuclear ^{1}H-^{15}N NOE values show low rigidity of both the PrgI N- and C-termini, suggesting the two termini are somewhat flexible, but

not in random coil conformation (Fig. 4-5). These data were collected at high protein concentration and whether less concentrated protein which excludes PrgI-PrgI interaction will show a lower order of rigidity is worthy of pursuit. The model of the *Shigella* needle suggests well-structured α-helices at the N- and C- termini of MxiH.[3] Nevertheless, these findings are consistent with the partial α-helical NOE pattern of PrgI and imply there might be a thermodynamically favored conformation changes upon needle assembly.

The C-terminal 5 residues of PrgI are important for the needle monomer-monomer interactions and needle formation.[1] It is not possible to use the $PrgI^{C\Delta5}$ construct to provide information on how C-terminal 5 residues are involved in the PrgI monomer-monomer interaction. Thus, a GB1 tag was used to solubilize the full length PrgI and both GB1-$PrgI^{FL}$ and $PrgI^{FL}$-GB1 were shown to be highly soluble in solution, suggesting that bulky GB1 tag prevented the needle protein from aggregation. Both GB1-$PrgI^{FL}$ and $PrgI^{FL}$-GB1 also showed concentration effect upon dilution, indicating that the GB1 tag might only disrupt some of the contacts that mediate the PrgI monomer-monomer interaction.

The chemical shift mapping data of ^{15}N-GB1-$PrgI^{FL}$ titrated with $PrgI^{C\Delta5}$ showed that C-terminal 5 residues did not lead to the same degree of peak shifts compared to the shifts seen for the head group residues (Fig. 4-6A and 6B). This suggests the C-terminal 5 residues might not play a dominant role in holding the needle monomers together as expected. Nevertheless, C-terminal 5 residues are still involved in one of the important contacts that glue the needle monomers together

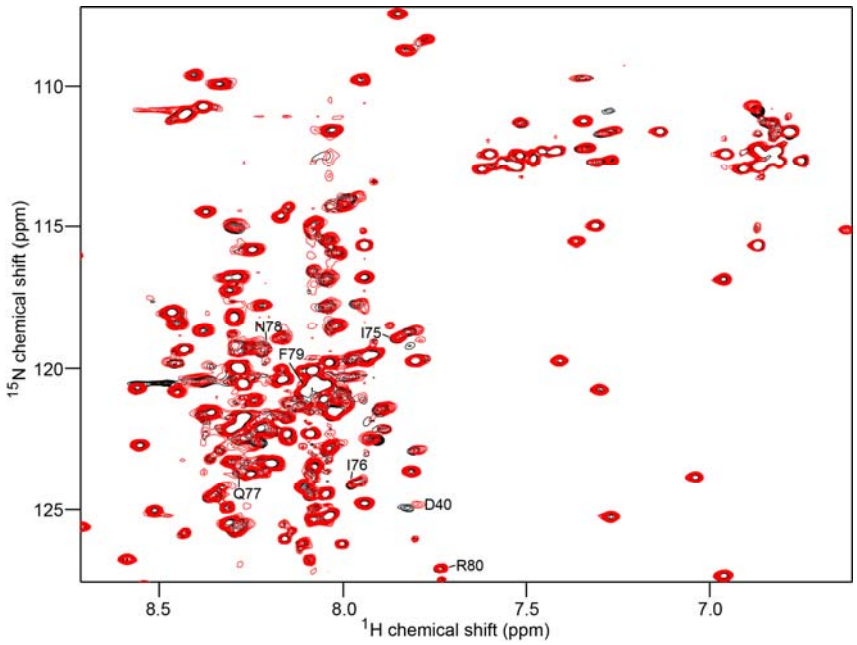

Figure 4-6A. Titration of ^{15}N-GB1-PrgI^{FL} with increasing amounts of unlabeled PrgI$^{C\Delta5}$. Overlay of two 2D HSQC spectra at various PrgI$^{C\Delta5}$:GB1-PrgI^{FL} molar ratios (*black*, 0 and *red*, 8.1). Clear shifts were observed upon addition of unlabeled SipD$^{N\Delta38}$ and peaks of C-terminal residues I75, I76, Q77, N78, F79 and R80 are pointed out to show relatively small or no change compared with residue D40.

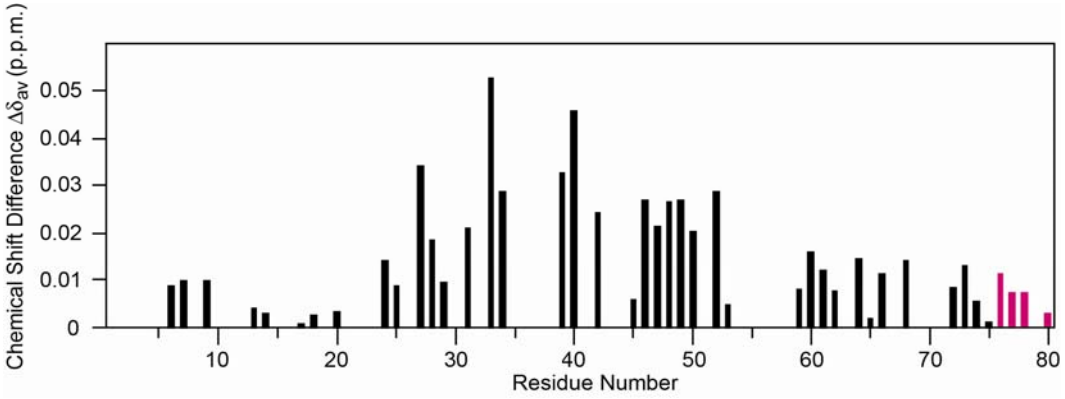

Figure 4-6B. Chemical shift differences for GB1-PrgI^{FL} upon addition of unlabeled PrgI$^{C\Delta5}$ comparing two different titration points at GB1-PrgI^{FL}: PrgI$^{C\Delta5}$ molar ratios of 0 and 8.1. The chemical shift differences were calculated as described. The X-axis is the residue number and the Y axis is the chemical shift differences $\Delta\delta_{av}$ (ppm). Only backbone amides are shown and calculated for chemical shift differences with the missing values being due to incomplete assignments or overlapped peaks. The last five residues (shown in red) were pointed out to show relatively small chemical shift difference.

among packed needle, suggesting the binding affinity among assembled needle monomers was dramatically decreased by the disruption of GB1 tag or C-terminal 5 residue deletion. We completed the backbone assignment of GB1-PrgI^{FL} including the C-terminal 5 residues. Further analysis of the C_α chemical shift difference between GB1-PrgI^{FL} and PrgI$^{C\Delta5}$ suggested GB1-PrgI^{FL} adopted essentially identical structures as PrgI$^{C\Delta5}$ (Fig. 4-S1A and S1B), implying that the C-terminal 5 residues are not sufficient in folding the N-terminal helix into a well formed helix by the interaction between two termini, and the N-terminal helix might rely on the inter-molecular interaction to become folded. It is likely that the two flexible termini of the needle protein will undergo conformational changes from free soluble protein into ordered assembled needle and this conformation change might be thermodynamically driven.

C-terminal 5 residue deletion is sufficient to prevent protein aggregation and needle assembly. It will be interesting to determine whether the N-terminus plays a similar role as the C-terminus. Thus, we made N-terminal deletion mutants, PrgI$^{N\Delta5}$, PrgI$^{N\Delta9}$, PrgI$^{N\Delta12}$, and PrgI$^{N\Delta16}$ and all purified proteins still slowly precipitated out of solution, implying the N-terminus might not be as important for needle packing compared to the C-terminal residues. It was shown, however, that residues Gly7 and Leu9 also have significant chemical shift changes in the titrations (Fig. 4-1B). These N terminal residues were also shown to have interactions with residues Lys66, Val67 and Phe68 close to C-terminus in the PrgI$^{C\Delta10}$ dilution data (Fig. 4-4B). The farthest long range ^{1}H-^{1}H NOEs from the PrgI two helices were between residues Phe16 and

Gln61. It is most likely that PrgI monomer-monomer interaction results in allosteric effects occurring at residues Gly7 and Leu9 through inter-molecular interactions. The N-terminal region might form an ordered α-helical structure upon needle assembly as proposed in the *Shigella* needle model.[3] The highly flexible residue Gly7 might play a role as a hinge that attaches the first 6 residues (MATPWS) to the rest of the molecule. The formation of α-helical structure might affect the inner lining of the needle channel. Although it is likely that the N-terminus is involved in protein-protein interaction between the needle channel and proteins passing through the channel, no experimental data are currently available showing the N-terminus is involved in protein-protein interactions.

The full length PrgI could be cleaved from purified GB1-PrgI^{FL} by TEV cleavage, but the protein slowly precipitated and formed gel. We were able to collect the supernatant after centrifugation, and the resulting 2D HSQC showed different spectra from that of soluble PrgI$^{C\Delta5}$, suggesting that full-length PrgI formed soluble aggreagates (Fig. 4-S5). Full length PrgI could also be purified in denaturing conditions (buffer that contains 6M urea). Once the urea was slowly dialyzed out, the protein formed white gel-like precipitate, especially when the urea concentration was lower than 1.0 M. Full length MxiH behaved similarly, however, full length BsaL is relatively soluble at low protein concentrations (0.1 mM) and 2D HSQC of full length BsaL shows essentially identical spectra compared to that of soluble BsaL$^{C\Delta5}$ (Fig. 4-S2A), suggesting a unique behavior for BsaL. This is consistent with the observation that BsaL$^{C\Delta5}$ does not have concentration effect compared to PrgI$^{C\Delta5}$ and MxiH$^{C\Delta5}$.

The N-terminal deletion construct $BsaL^{N\Delta25}$ also behaved differently compared to the PrgI N-terminal deletion construct, $PrgI^{N\Delta16}$; and $BsaL^{N\Delta25}$ is highly soluble and monomeric in solution (Fig. 3-S8). These differences probably suggest BsaL might assemble differently with PrgI and MxiH.

In order to probe how needles are packed together, further NMR characterizations including identification of distance restraints between needle monomers using NOE, RDC (residual dipolar coupling) and PRE (paramagnetic relaxation enchancement) restraints are needed. High resolution techniques like cryo-EM or solid state NMR will also provide important information on the self-associated high molecular weight protein complex. Additionally, *in vivo* experiments to test the biological relevance of structural data are needed to achieve a better understanding of the needle assembly. We are going to perform *Salmonella* invasion assays on PrgI mutants and the deficiency of invasiveness will be further investigated by immunofluorescence staining of surface-exposed bacterial needle and transmission electron microscopy of the sheared needles from nonflagellated *ΔflhC Salmonella typhimurium*. The protein expression level and the secretion of proteins SipB and SipC into the culture supernatant will be tested by western-blot analysis.

REFERENCES

1. Kenjale, R., Wilson, J., Zenk, S. F., Saurya, S., Picking, W. L., Picking, W. D. & Blocker, A. (2005). The needle component of the type III secreton of *Shigella* regulates the activity of the secretion apparatus. *J Biol Chem* **280**, 42929-37.

2. Davis, A. J. & Mecsas, J. (2007). Mutations in the *Yersinia pseudotuberculosis* type III secretion system needle protein, YscF, that specifically abrogate effector translocation into host cells. *J Bacteriol* **189**, 83-97.
3. Deane, J. E., Roversi, P., Cordes, F. S., Johnson, S., Kenjale, R., Daniell, S., Booy, F., Picking, W. D., Picking, W. L., Blocker, A. J. & Lea, S. M. (2006). Molecular model of a type III secretion system needle: Implications for host-cell sensing. *Proc Natl Acad Sci U S A* **103**, 12529-33.
4. Wang, Y., Ouellette, A. N., Egan, C. W., Rathinavelan, T., Im, W. & De Guzman, R. N. (2007). Differences in the electrostatic surfaces of the type III secretion needle proteins PrgI, BsaL, and MxiH. *J Mol Biol* **371**, 1304-14.
5. Zhang, L., Wang, Y., Picking, W. L., Picking, W. D. & De Guzman, R. N. (2006). Solution structure of monomeric BsaL, the type III secretion needle protein of *Burkholderia pseudomallei*. *J Mol Biol* **359**, 322-30.
6. Wang, Y., Zhang, L., Picking, W. L., Picking, W. D. & Guzman, R. N. D. (2008). Structural dissection of the extracellular moieties of the type III secretion apparatus. *Mol. BioSyst.* **4**, 1176-80.
7. Cordes, F. S., Komoriya, K., Larquet, E., Yang, S., Egelman, E. H., Blocker, A. & Lea, S. M. (2003). Helical structure of the needle of the type III secretion system of *Shigella flexneri*. *J Biol Chem* **278**, 17103-7.
8. Grzesiek, S., Dobeli, H., Gentz, R., Garotta, G., Labhardt, A. M. & Bax, A. (1992). 1H, 13C, and 15N NMR backbone assignments and secondary structure of human interferon-gamma. *Biochemistry* **31**, 8180-90.
9. Kay, L. E., Torchia, D. A. & Bax, A. (1989). Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972-9.
10. Grzesiek, S. & Bax, A. (1993). The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115**, 12953-12954.
11. Stone, M. J., Fairbrother, W. J., Palmer, A. G., 3rd, Reizer, J., Saier, M. H., Jr. & Wright, P. E. (1992). Backbone dynamics of the Bacillus subtilis glucose permease IIA domain determined from 15N NMR relaxation measurements. *Biochemistry* **31**, 4394-406.
12. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-93.
13. Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* **278**, 313-52.
14. Tang, C., Iwahara, J. & Clore, G. M. (2006). Visualization of transient encounter complexes in protein-protein association. *Nature* **444**, 383-6.
15. Coudevylle, N., Montaville, P., Leonov, A., Zweckstetter, M. & Becker, S. (2008). Structural determinants for Ca2+ and phosphatidylinositol 4,5-bisphosphate binding by the C2A domain of rabphilin-3A. *J Biol Chem* **283**, 35918-28.

SUPPLEMENTAL FIGURES

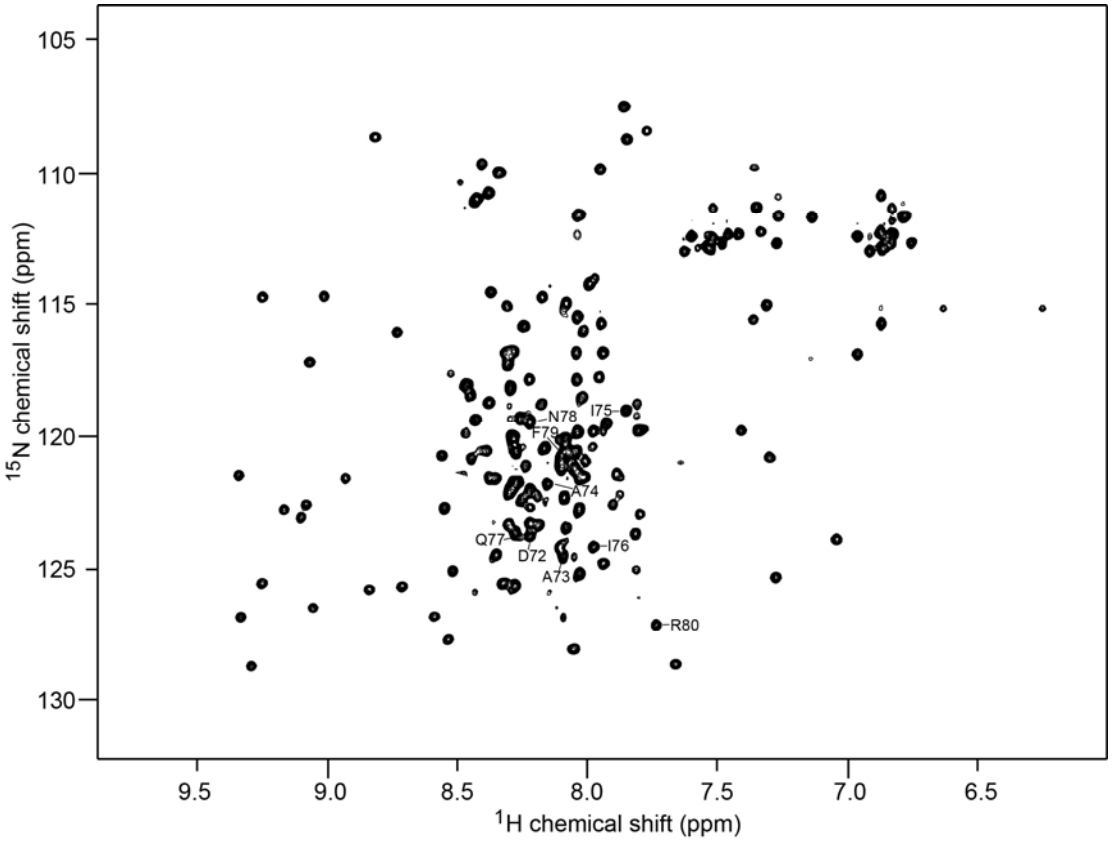

Figure 4-S1A. 2D HSQC spectra of GB1-PrgI^{FL} with sequence specific assignments of the last 10 residues. The assignments were finished based on 3D experiments: 3D HNCA and CBCA(CO)NH spectra using ^{13}C, ^{15}N-labeled GB1-PrgI^{FL}.

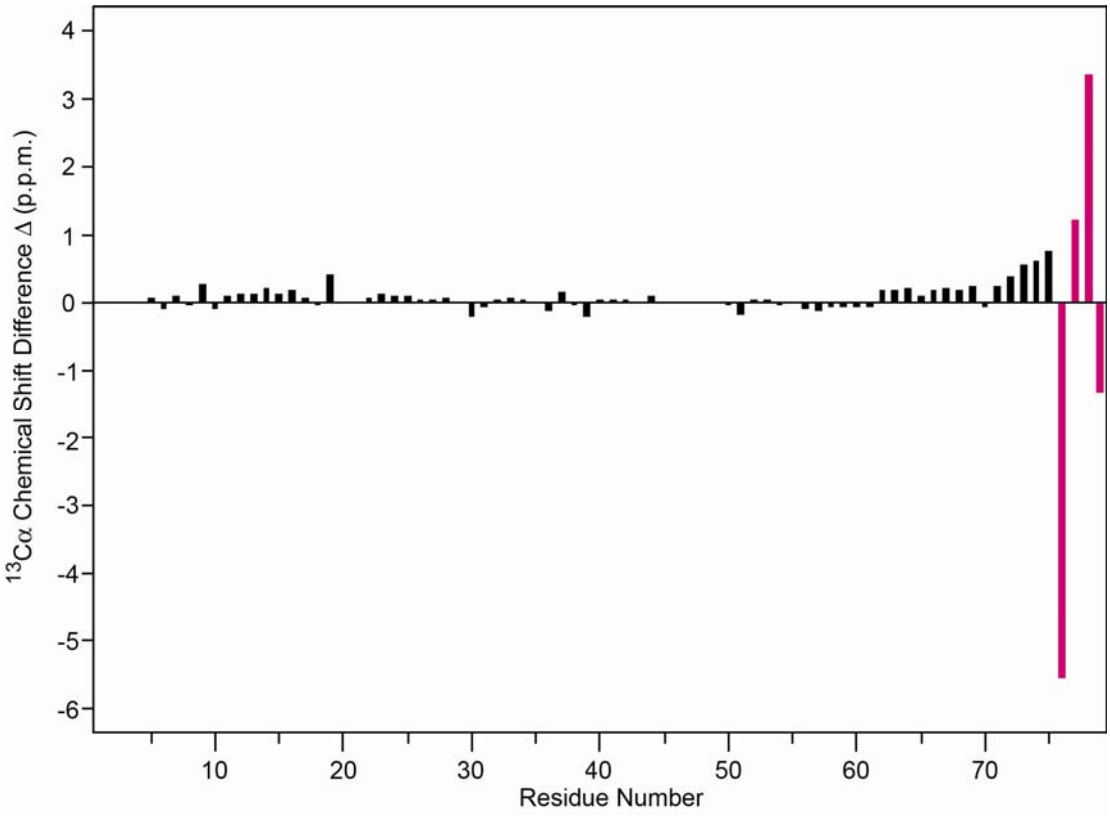

Figure 4-S1B. Difference between the C_α chemical shifts in $PrgI^{C\Delta5}$ and C_α chemical shifts in GB1-$PrgI^{FL}$. Nearly no difference for the $PrgI^{C\Delta5}$ residues (black bars) was observed, indicating similar structure in both constructs $PrgI^{C\Delta5}$ and GB1-$PrgI^{FL}$. Only a few residues close to the C-terminal showed relatively higher chemical shift differences. The high chemical shift values (red bars) are due to the comparison of different residues in the C-terminal residues.

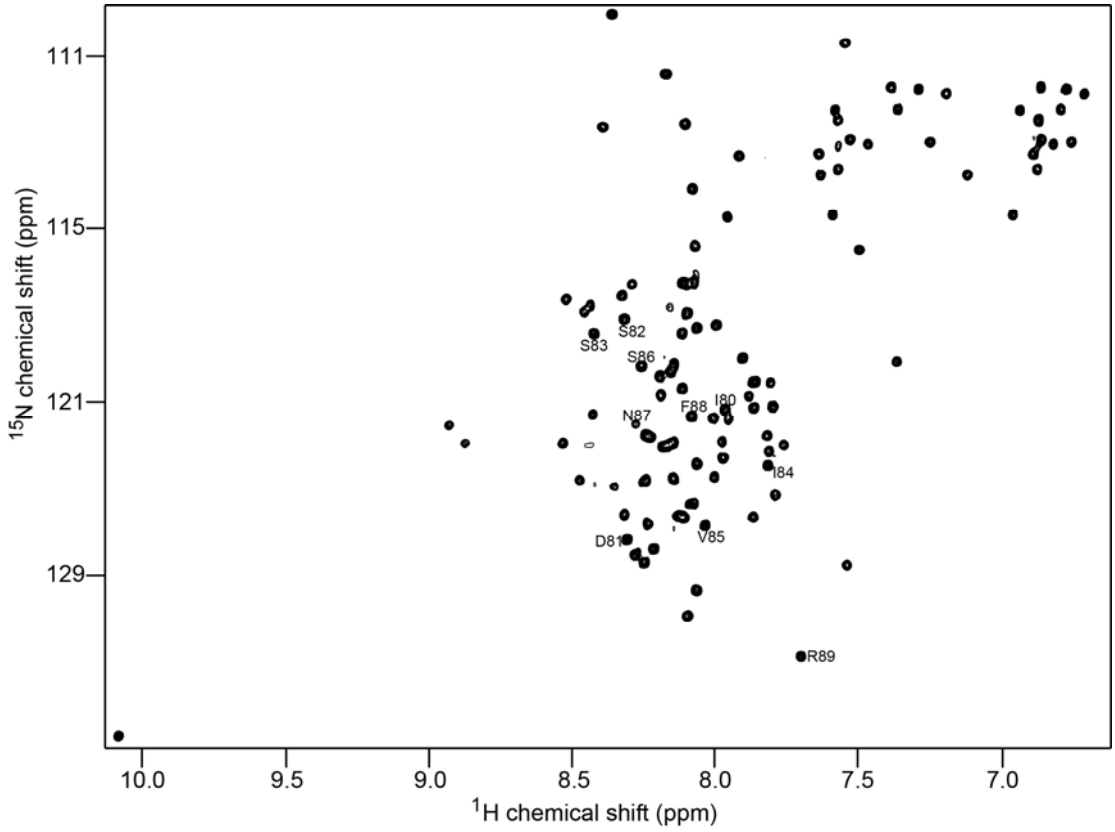

Figure 4-S2A. 2D HSQC spectra of BsaL^{FL} with sequence specific assignments of the last 10 residues. The assignments were finished based on 3D experiments: 3D HNCA and CBCA(CO)NH spectra using ^{13}C, ^{15}N-labeled GB1-BsaL^{FL}.

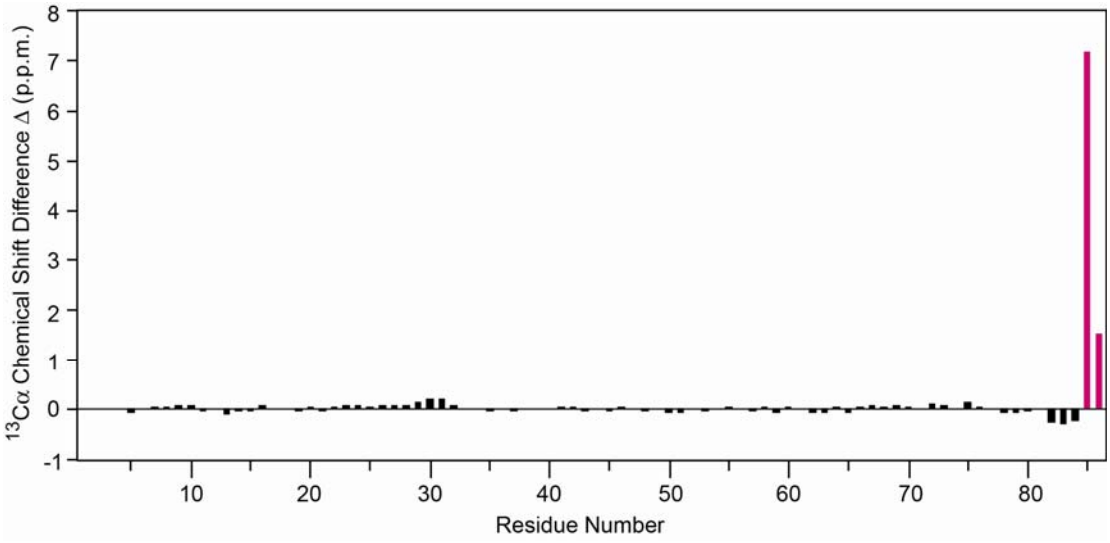

Figure 4-S2B. Difference between the C_α chemical shifts in GB1-BsaL^{FL} and C_α chemical shifts in BsaL$^{C\Delta5}$. Nearly no difference for the BsaL$^{C\Delta5}$ residues (black bars) was observed, indicating similar structure between BsaL$^{C\Delta5}$ and GB1-BsaL^{FL}. The high chemical shift values (red bars) are due to the comparison of different residues in the C-terminal residues.

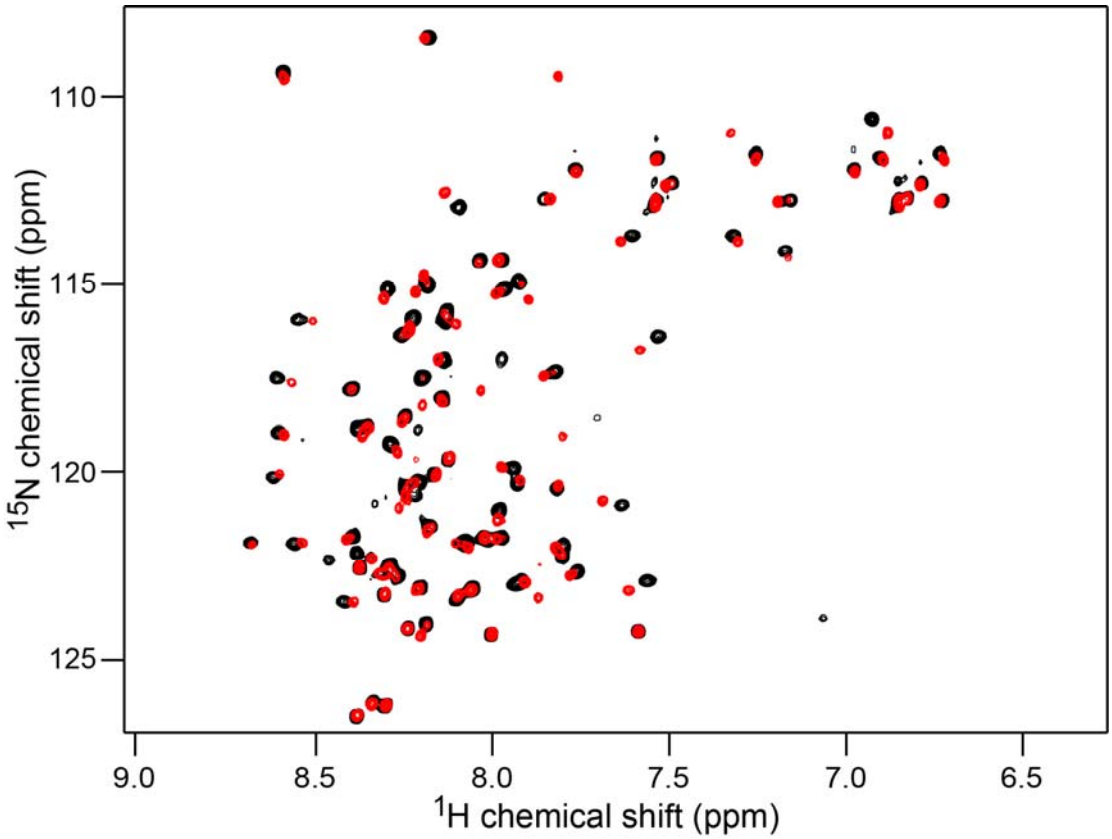

Figure 4-S3. Dilution of ^{15}N-MxiH$^{C\Delta5}$ with increasing amounts of NMR buffer (10 mM Sodium Phosphate pH 6.5). Overlay of two 2D HSQC spectra at various dilution factors (*black*, 0 and *red*, 10). Some of MxiH$^{C\Delta5}$ residues showed obvious peak shifts.

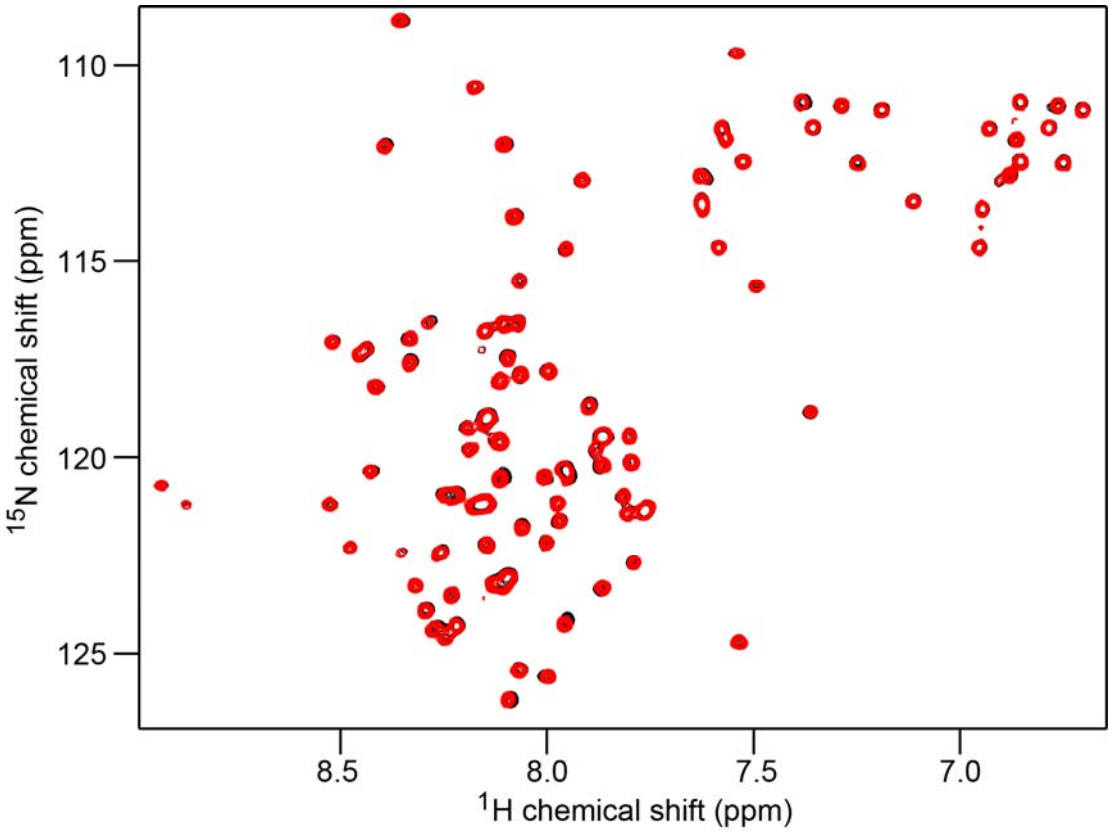

Figure 4-S4. Dilution of ^{15}N-BsaL$^{C\Delta5}$ with increasing amounts of NMR buffer (10 mM Sodium Phosphate pH 6.5). Overlay of two 2D HSQC spectra at various dilution factors (*black*, 0 and *red*, 10). No obvious peak shifts were observed upon dilution.

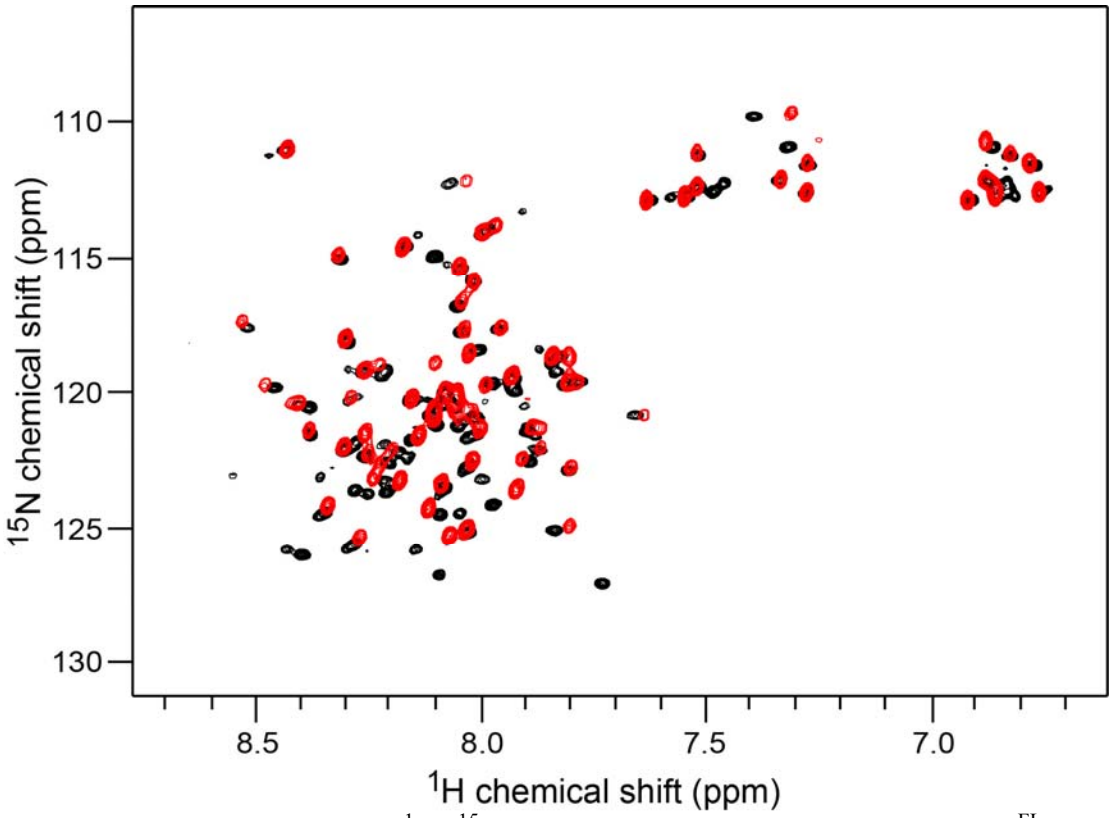

Figure 4-S5. Overlay of 2D ^{1}H, ^{15}N-HSQC spectra of supernatant of PrgI^{FL} gel-forming sample (*Black*) with soluble PrgI$^{C\Delta5}$ (*Red*). Lots of the PrgI$^{C\Delta5}$ peaks did not overlay completely, suggesting different PrgI structural forms.

Chapter 5: Characterization of Needle-Tip (PrgI-SipD) Interaction

INTRODUCTION OF TIP COMPLEX

It was not until late 2005 that a TTSA tip complex composed of LcrV was reported at the tip of the *Yersinia enterocolitica* needle when the apparatus was in a resting state.[1] This was quickly supported by showing that IpaD formed a tip complex for TTSAs in *S. flexneri*.[2] It is proposed that the tip complex is needed to sense environmental signals and control the secretion and insertion of translocators into host membranes to initiate translocon pore formation[3; 4; 5] In *Shigella*, once both translocator proteins IpaB and IpaC are recruited and inserted into the host cell membrane, a pore-like structure called the translocon is formed. Different effector proteins through the translocon pore are further injected directly into the host cell cytosol to disrupt normal cellular functions. Using atomic force microscopy, the translocon pore size of the translocon from different species was shown to vary from 1.5 to 3.5 nm.[6] Overall, the tip complex is believed to serve as a platform for the assembly of the translocon, which ultimately connects the bacterial pathogen to its eukaryotic host for the delivery of effector proteins.[7]

Deletion of the first 20 residues of the IpaD prevents its secretion, suggesting that these residues act as a secretion signal.[3] Deletions within the N-terminal third of the molecule do not affect invasion of host cells by the bacteria, although the insertion of IpaB/IpaC into membranes is slightly impaired.[3] However, even short deletions (5 residues) in the C terminus of IpaD completely abolish the invasive

phenotype and pore insertion.[3] These observations suggested that IpaD may play a role in regulating insertion of the IpaB/IpaC translocon pore from the tip of the *S. flexneri* needle.[3] Sequence similarity suggests that SipD and BipD might play a similar role as IpaD.[8] However, the effect of SipD-dependent control of effector secretion on *Salmonella* invasion has yet to be clearly described.

Based on the sequence similarity, especially within their C-terminal sequences, tip proteins have been distinguished into two specific classes: (a) those of the SipD/IpaD/BipD family, and (b) those of the LcrV (*Yersinia* spp.)/PcrV (*Pseudomonas aeruginosa*) family.[9] SipD, IpaD and BipD share ~30% primary sequence similarity and more than 90% similarity at the C-termini (Fig. 5-1), and very little with LcrV/PcrV. Biophysical data showed that all tip proteins possess the α-helical properties, which is consistent with the presence of a core coiled-coil in the X-ray structures (Fig. 5-2).[9] Both SipD and IpaD showed two thermal transitions in CD denaturation experiments, suggesting the presence of two independently folded domains.[9] This unique feature implicates distinct function of SipD and IpaD from other tip proteins.[9]

The first reported high resolution structure of tip protein is the X-ray structure of monomeric LcrV (Fig. 5-2).[10] LcrV has a dumbbell shape with a long central coiled-coil flanked by two globular domains.[10] The long central coiled-coil was found to be a distinguishing feature of other tip proteins when the X-ray structures of BipD[5; 11] and IpaD[5] became available. Both BipD and IpaD, however, possess larger N-terminal domains than LcrV (Fig. 5-2). It is perhaps not surprising that the N-terminal

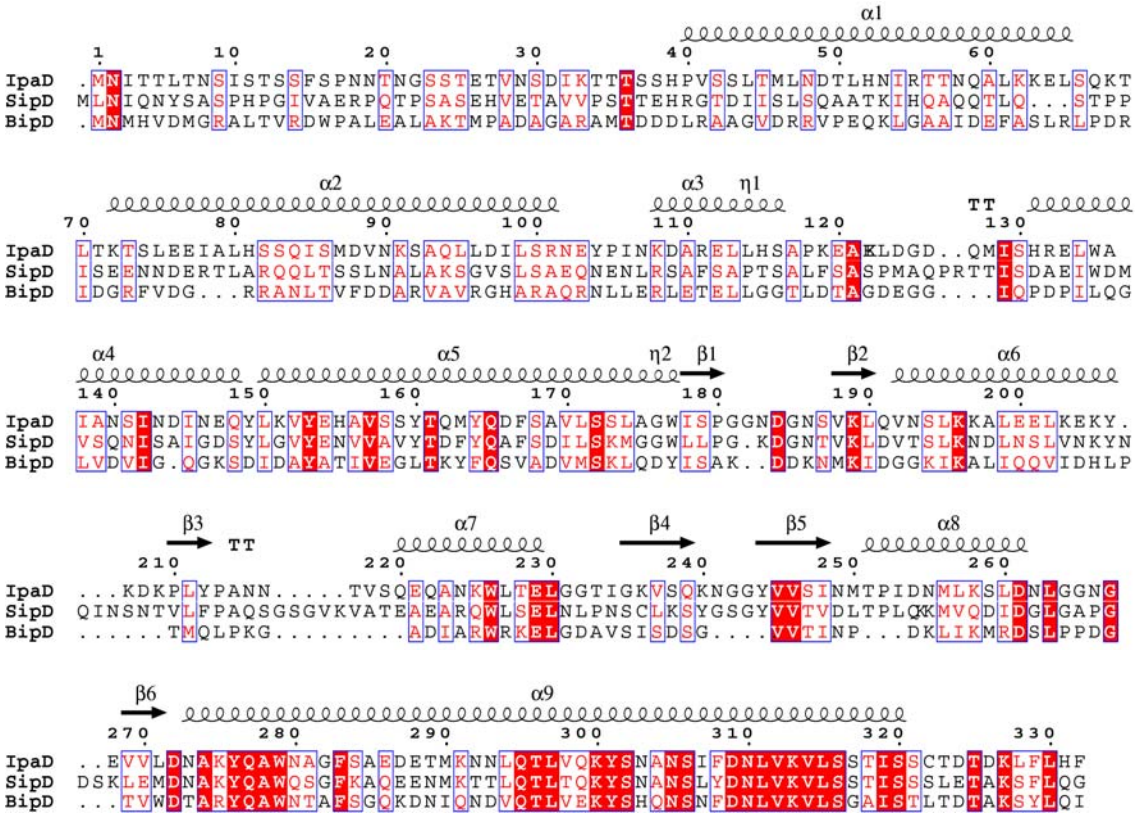

Figure 5-1. TTSS tip proteins of different bacteria show sequence conservation, especially within the C-terminal region. Highly conserved residues are colored red with the stretch conserved residues are boxed and identical residues are shaded red. The assigned secondary structure on top of the sequences is based on the crystal structure of IpaD. Sequences are from Swiss-Prot, and the bacterial species and its associated needle protein are: *S. flexneri* (IpaD), *S. typhimurium* (SipD) and *B. pseudomallei* (BipD).

domains of BipD and IpaD resemble the structure of the LcrV chaperonee LcrG since homologs of this chaperonee in *Shigella* and *Burkholderia* are absent and it has been suggested that IpaD and BipD may possess a self-chaperoneing ability.[5]

Although the tip complex has been visualized at the tip of the needle,[1; 2] how the tip protein assembles is not known. Based on immunoblotting and scanning-transmission electron microscopy (STEM), it was suggested that three to five LcrV molecules were needed to form the tip complex of the *Yersinia* needle.[12] Molecular modeling based on the crystal structures of LcrV and *Shigella* needle model suggested a pentameric LcrV tip complex.[13] Likewise, molecular modeling based on the crystal structures of *Shigella* tip protein IpaD and its needle model suggested a pentameric IpaD tip complex.[5] The tip proteins have not, however, been found to self-assemble into pentamers *in vitro* during structure analyses which makes solving the the native tip complex more challenging. How the tip proteins interact with each other also remains to be determined. Caroline *et al*. were able to unfold the tip protein PcrV and LcrV by lowering the pH and the oligomeric species of tip proteins formed after refolding.[14] However, key information about the difference between self-association form and monomeric form of tip proteins is absent.

The positioning of the tip complex suggests that it interacts directly with the needle. Precisely how the tip proteins interact with needle monomers is currently unknown. The N-terminal domain of the IpaD tip protein has a two helix bundle that looks remarkably similar to the two helix bundle of the MxiH needle protein, suggesting a possible model for MxiH–IpaD interaction.[15] A model suggests that

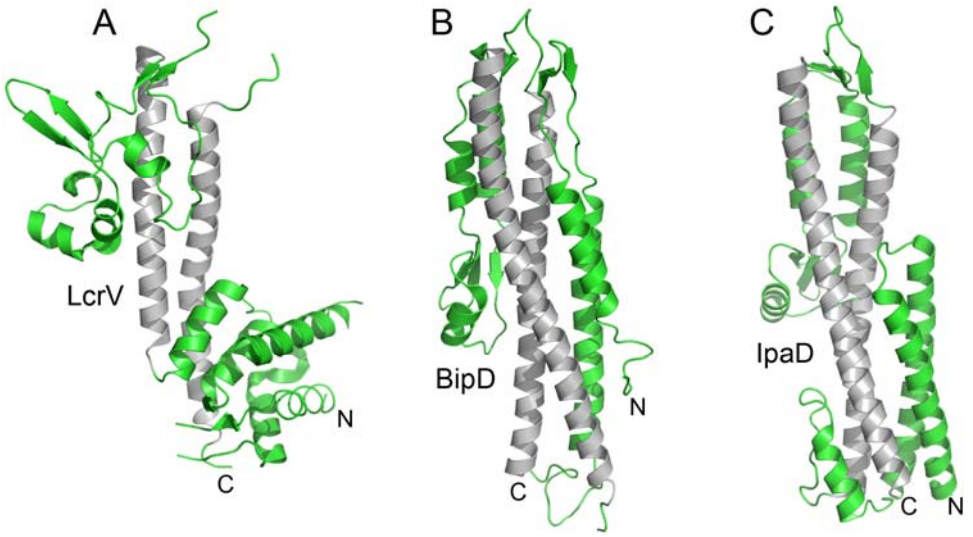

Figure 5-2. Ribbon structures of tip proteins (**A**) LcrV, (**B**) BipD, and (**C**) IpaD. Figures were generated using the following PDB IDs: LcrV, 1R6F; BipD, 2IZP; and IpaD, 2J0O.[7] The central coiled-coil (gray) was flanked by two separate N-terminal and distal domains (green).

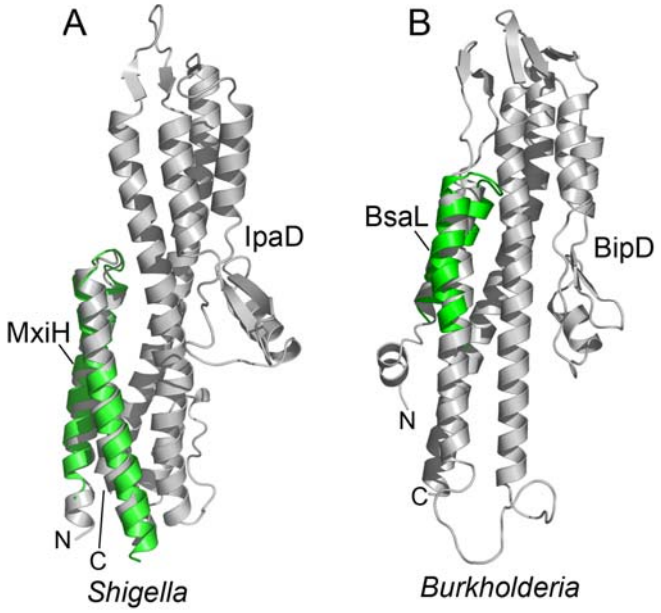

Figure 5-3. A model of MxiH–IpaD interaction is based on the structural similarity of these proteins. (**A**) The MxiH two helix bundle (green) can be superimposed with the N- terminal domain of IpaD (gray) with a C_α root-mean-square deviation of 3.2 Å. (**B**) Likewise, the BsaL two helix bundle (green) can be superimposed with the N-terminal domain of BipD.[7]

when IpaD is secreted and localized to the tip of the needle, its N-terminal domain swings away, leaving the binding interface of the long central coiled-coil accessible to MxiH.[15] MxiH thus mimics the IpaD N-terminus and binds to the lower part of the IpaD long central coiled-coil (Fig. 5-3). It is likely that other tip proteins will bind their respective needle partners in a similar manner based on the following observations: (i) the two helix bundle of BsaL can be fitted with the N-terminal two helix bundle of BipD (Fig. 5-3); (ii) BipD is also capable of binding the MxiH needle of *S. flexneri*,[5] and the C-terminal sequences of IpaD and BipD are similar. Likewise, the divergence of C-terminal sequences for the LcrV/PcrV family and their inability to be retained at the *Shigella* MxiH needle tip (unpublished data) are also consistent with this model of MxiH–IpaD interaction (Fig. 5-3).

In this chapter, we show the characterization of the needle-tip interaction using NMR. We mainly focused on the interactions between PrgI and SipD (from *Salmonella*) and BsaL and BipD (from *Burkholderia*). Our titration data showed distinct binding sites of PrgI are involved in the PrgI-SipD interaction compared to the PrgI-PrgI interaction. It was also suggested that the interactions between the needle proteins and tip proteins are weak. Detailed understanding of the needle-tip interactions requires further determination of the high resolution structures of these complexes.

Part I: Characterization of PrgI bound to SipD

INTRODUCTION

The *Shigella* tip protein IpaD localizes on top of the *Shigella* needle and controls the secretion and insertion of translocator proteins IpaB and IpaC into the host membrane.[2; 3] The tip protein SipD (*Salmonella* invasion plasmid antigen D) is likely to play a similar role. SipD was shown to be crucial for protein secretion and *Salmonella* invasion.[8] It was suggested that tip proteins assemble on top of the needle and stabilized by the interaction between tip proteins and needle proteins.[1; 13] A high resolution structure of SipD is unavailable; however, SipD shares high sequence conservation with IpaD whose structure is known.[5] Using the structure prediction program I-TASSER, SipD is suggested to have a similar structure to IpaD. We solved the structure of the *Salmonella* needle protein PrgI by NMR and PrgI was shown to possess a helix-turn-helix structure with a central PxxP motif.[16] The N- and C-termini flanking the core structure adopt partial helical structures. The individual high-resolution structures of IpaD, PrgI and the modeled structure of SipD have provided valuable information in understanding the function of these proteins. However, questions about the assembly mechanism of type III secretion tip complex at the needle tip still remain to be answered.

Here, we describe the characterization of how the *Salmonella* PrgI (needle protein) interact with the SipD (tip protein). NMR chemical shift mapping was used to show that distinct PrgI residues were clustered and involved in the PrgI-SipD

interaction. The PrgI residues involved in the PrgI-SipD interaction are clustered on a large patch close to the C-terminus of PrgI. In addition, a couple of residues at the N-terminus are also involved .Our data suggest that different subsets of PrgI residues are involved in needle assembly (PrgI-PrgI) and needle-tip (PrgI-SipD) interaction. Although the main focus in this chapter is the needle (PrgI) and tip (SipD) proteins of *Salmonella*, we also obtained data for the corresponding proteins for *Burkholderia* needle (BsaL) and tip (BipD) proteins to serve as experimental control.

MATERIALS AND METHODS

Protein Expression and Purification

The SipD construct, $SipD^{N\Delta 38}$, was designed based on the sequence homology between SipD and IpaD that showed the first 38 residues of SipD are disordered. In the crystal structure of IpaD, the N-terminal 38 residues of IpaD lacked of electron density and were essentially unstructured. The $SipD^{N\Delta 38}$ gene was subcloned into the pET-21a (Novagen), which appends a His_6 tag followed by a GB1 domain and a TEV protease cleavage site at the amino terminus. The BipD construct, $BipD^{35-301}$, was designed based on its crystal structure (Fig. 5-1) which showed both terminal ends were disordered and thus were removed. The $BipD^{35-301}$ gene was subcloned into pET-21a (Novagen), which also appends a His_6 tag and a TEV protease cleavage site at the amino terminus. Isotopically (^{15}N) labeled protein was overexpressed in *Escherichia coli* BL21 (DE3) (DNAY) grown in 1 liter of M9 minimal medium with ^{15}N ammonium chloride (Cambridge Isotope Laboratories, Inc.). The bacteria were

grown at 37˚C until mid-log phase, induced with 1 mM IPTG and incubated in 15˚C shaker overnight. The bacterial cells were harvested, resuspended in 30 ml of binding buffer (20 mM Tris.HCl (pH 7.9), 500 mM NaCl and 5 mM imidazole) and lysed by sonication. The recombinant proteins were purified by Ni^{2+} affinity chromatography as described.[16] The His_6-GB1 tag or His_6 tag was removed by TEV protease cleavage from purified His_6-GB1-(TEV)-$SipD^{N\Delta 38}$ and His_6-(TEV)-$BipD^{35-301}$ respectively as described.[17] The TEV protease cleaved proteins were further purified by passing through Ni^{2+} affinity column. All purified protein samples were concentrated using Amicon ultracel-3k when needed and dialyzed into NMR buffer (10 mM sodium phosphate, pH 6.5, 10 mM NaCl) before use, and the protein concentration was estimated using absorbance at 280 nm.

NMR Spectroscopy

Uniformly ^{15}N-labeled $PrgI^{C\Delta 5}$ and unlabeled $SipD^{N\Delta 38}$ were dialyzed in NMR buffer (10 mM sodium phosphate pH 6.5, 10 mM NaCl). NMR data were acquired at 25°C using a Bruker Avance 800 MHz spectrometer equipped with a cryogenic Z-axis pulse field gradient probe. NMR chemical shift mapping was done by acquiring a series of ^{1}H-^{15}N HSQC 2D spectra. NMR data were processed using NMRPipe [18] and analyzed using NMRView [19].

NMR Chemical Shift Mapping

Purified recombinant ^{15}N-PrgI$^{C\Delta5}$ and unlabeled SipD$^{N\Delta38}$ were purified as described and the concentration is estimated by UV absorbance to be 280nm as 0.97 mM and 0.72 mM, respectively. To eliminate the concentration effect, we perform the chemical shift mapping titrations using fixed amounts of ^{15}N- PrgI$^{C\Delta5}$ and mixing with various amount of unlabeled protein GB1-PrgI^{FL} and SipD$^{N\Delta38}$ in different NMR tubes by compensating different amount of buffer with different protein volume ratios of 1:1, 1:3, 1:6 and 1:11. ^{15}N-PrgI$^{C\Delta5}$ dilution experiment was done by diluting the protein into different dilution factors of 0, 3, 5 and 12. Purified protein ^{15}N-PrgI$^{C\Delta5}$ and SipD$^{N\Delta38}$ were dialyzed into the same buffer prior to titration to assure that the chemical shift changes were only due to protein-protein interactions and not to changes in solution condition upon mixing.

RESULTS

PrgI$^{C\Delta5}$ Residues Perturbed in PrgI-SipD Titration

The concentration of ^{15}N-labeled PrgI$^{C\Delta5}$ was kept identical in all titration points to eliminate the concentration effect during the titration of ^{15}N-labeled PrgI$^{C\Delta5}$ with unlabeled SipD$^{N\Delta38}$ (Fig. 5-4A). Peak shifts were observed, but the chemical shift changes were smaller than the changes observed, albeit smaller than the chemical shift changes observed in the titrations for PrgI monomer-monomer interaction (Fig. 5-4A). The small chemical shift changes indicate weaker interaction between PrgI and SipD in the order of 1.0 mM, and the interaction is in fast exchange NMR time scale. The direction of the peaks shifts is in a different direction with

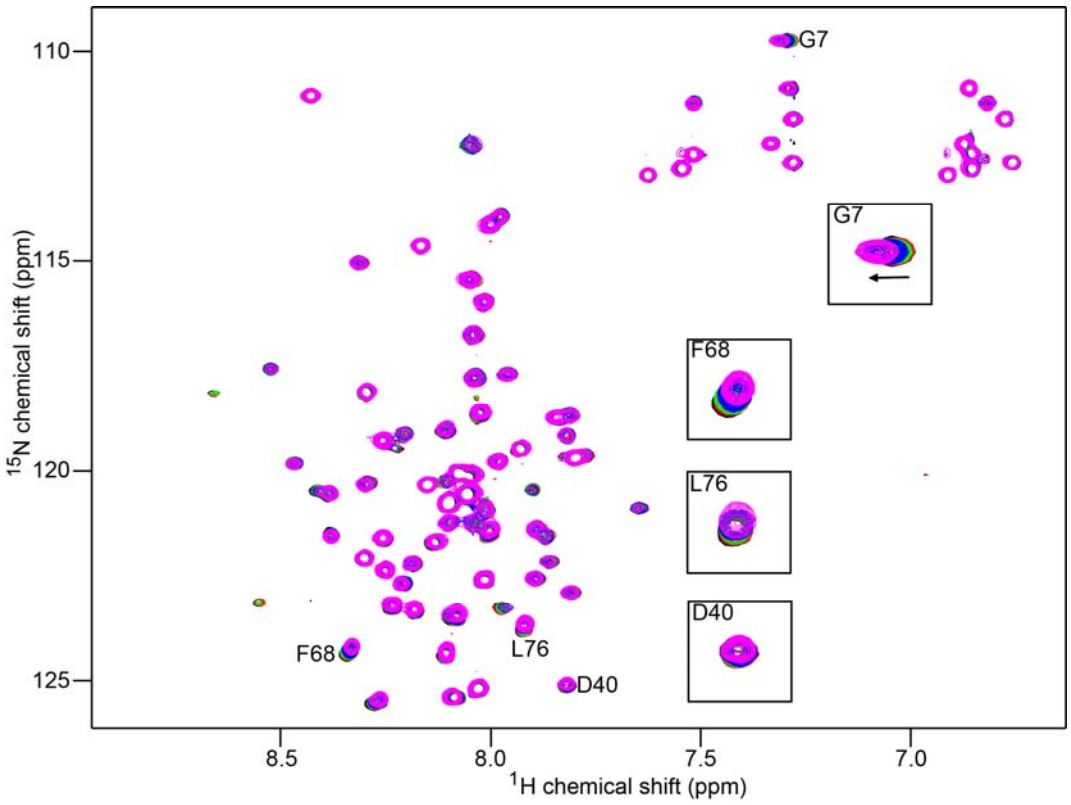

Figure 5-4A. Titration of ^{15}N-PrgI$^{C\Delta5}$ with increasing amounts of unlabeled SipD$^{N\Delta38}$. Overlay of five 2D HSQC spectra at various SipD$^{N\Delta38}$:PrgI$^{C\Delta5}$ molar ratios (*black*, 0; *red*, 0.74; *green*, 2.23; *blue*, 4.45; *magenta*, 8.16). Peak shifts were observed upon addition of unlabeled SipD$^{N\Delta38}$. Peaks of specific residues G7, F68 and L76 that showed peak shifts are boxed and zoomed in (2x) for clarity. D40, which did not show peak shifts in the PrgI-SipD titration, but showed a significant peak shifts in the PrgI-PrgI titrations (Fig. 4-1A) is also boxed.

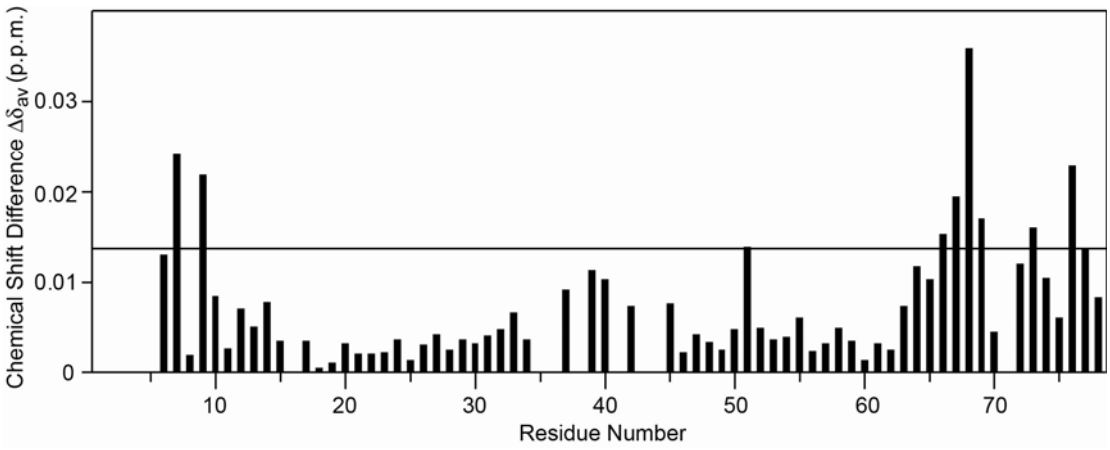

Figure 5-4B. Chemical shift differences for PrgI$^{C\Delta5}$ upon addition of unlabeled SipD$^{N\Delta38}$ at two titration points with SipD$^{N\Delta38}$: PrgI$^{C\Delta5}$molar ratios of 0 and 8.16. The chemical shift differences were calculated according to the formula $\Delta\delta_{av} = \sqrt{((\Delta\delta_{HN})^2/2 + (\Delta\delta_N/5)^2/2)}$, where $\Delta\delta_{HN}$ and $\Delta\delta_N$ are the amide proton and nitrogen chemical shift differences, respectively. The x-axis is the residue number and the y axis is the chemical shift differences $\Delta\delta_{av}$ (ppm). Only backbone amides are shown. Missing values are due to incomplete assignments or overlapped peaks.

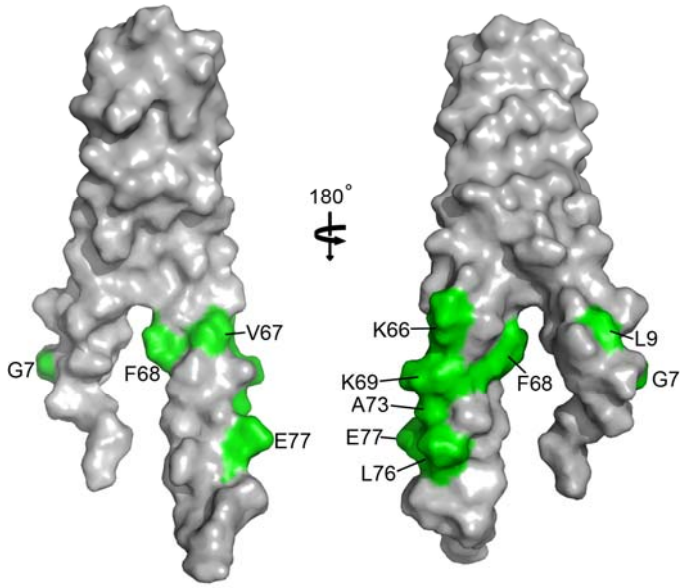

Figure 5-4C. Surface representation of the residues of PrgI$^{C\Delta5}$ (shown in green) that are perturbed upon PrgI-SipD interaction. The residues mainly cluster at the PrgI C-terminal tail. Two residues at the PrgI N-terminal are also perturbed.

peaks shifts observed upon the dilution of $PrgI^{C\Delta5}$ (Fig. 4-1A), indicating those chemical shift changes are due to the binding of SipD. Additionally, $PrgI^{C\Delta5}$ peak broadening was observed upon addition of unlabeled $SipD^{N\Delta38}$, suggesting that protein-protein interaction was occurring. Similar to the PrgI monomer-monomer interaction, the N-terminal residues Gly7 and Leu9 also showed chemical shift changes in the PrgI-SipD titration (Fig. 5-4B and 4C). The C-terminal residues Lys66, Val67, Phe68, Lys69 and residues Asp72, Ala73, Leu76 and Glu77 of $PrgI^{C\Delta5}$ showed relatively significant chemical shift changes and these chemical changes are different from those seen in the PrgI-PrgI interaction (Fig. 5-4B and 4C).

DISCUSSION

The NMR chemical shift mapping data of PrgI-SipD titration show that distinct binding sites are involved in needle-tip interaction compared with the PrgI-PrgI interaction. A stretch of residues close to the C-terminus, from Lys66 to Lys69 and from Asp72 to Glu77, showed chemical shift changes upon PrgI-SipD binding, suggesting these residues are most likely affected by the PrgI-SipD interaction. Among these residues, Lys66, Val67, Phe68 were also shown to be involved in PrgI needle monomer-monomer interaction (Fig. 4-1B). N-terminal residues Gly7 and Leu9 also showed chemical shift changes, suggesting that the PrgI-SipD interaction also requires the PrgI N-terminus. Alternatively, it is also possible that the N-terminus is not the direct binding site and that the N-terminal residues (Gly7 and Leu9) experience allosteric changes upon the PrgI-SipD binding. The induced

conformational changes at the N-terminal region might affect the function of the several N-terminal PrgI residues, which might be important for the protein-protein interactions occurring between the needle protein and proteins traveling through the needle channel. A possible scenario in the binding interaction would be a change in the orientation of N-terminus of the needle protein which lines the inner needle channel, where the invasion signal might be transmitted from the tip to the bacterial basal body. Nevertheless, our chemical shift mapping data showed that unlike the needle monomer-monomer (PrgI-PrgI) interaction, the needle-tip (PrgI-SipD) interaction mainly involves the C-terminus of PrgI and probably the N-terminus of PrgI as well. To have a better understanding of the PrgI-SipD interaction, attempts to characterize how SipD binds to PrgI are still needed.

Regarding the needle and tip proteins of *Burkholderia*, titration of ^{15}N-BsaL$^{C\Delta5}$ with unlabeled BipD$^{35\text{-}301}$ showed no definitive binding interface of BsaL-BipD interaction. The NMR titration data is consistent with the observation that interaction between PrgI and SipD are weak. Due to the structural role of both tip and needle proteins on the bacterial surface, it is possible that the local concentration of two proteins is high enough to compensate for the low binding affinity observed *in vitro* for isolated proteins. Therefore, a strong binding is not necessary and a fine regulation can be achieved by subtle structural changes upon binding. To achieve a better understanding of needle-tip interaction, PRE would be a useful alternative in studies of weak protein-protein interactions.

Part II: Characterization of Tip Protein Bound to Needle Protein

INTRODUCTION

After determining the binding interface of PrgI bound to SipD, we next addressed the question of how SipD binds to the needle protein. Compared to the relatively small needle proteins, the high molecular weight tip proteins pose additional challenges for NMR studies due to line broadening. Nevertheless, TROSY (Transverse Relaxation Optimized Spectroscopy), perdeuteration and specific labeling allowed the NMR resonance assignments of SipD.

Here, we used NMR chemical shift mapping to characterize how the tip (SipD) protein interacts with the needle (PrgI) protein. The chemical shift mapping of SipD upon titration of PrgI revealed no conclusive peak shifts, suggesting weak interaction between SipD and PrgI. However, chemical shift changes were observed for Burkholderia tip (BipD) and needle (BsaL) proteins, despite similar weak interaction between BipD and BsaL. The key residues of BipD involved in binding BsaL will be revealed upon the completion of the sequence specific assignment of BipD.

MATERIALS AND METHODS

Protein Expression and Purification

Perdeuterated ^{15}N, ^{13}C-labeled $SipD^{N\Delta 38}$ and $BipD^{35\text{-}301}$ samples were obtained by growing *Escherichia coli* BL21(DE3) in 1 liter of minimal media dissolved in

99.9% D_2O and supplemented with 2 g/L of [^{13}C]glucose and 1 g/L of [^{15}N]ammonium chloride. The perdeuterated ^{15}N, ^{13}C-labeled growth was ~ 2 hours slower than normal ^{15}N, ^{13}C-labeled growth before the bacterial density reached A_{600}~0.8. After induction, cell growth was continued at 15°C overnight and the A_{600} barely reached ~2.0. Isotopically (^{15}N) labeled SipD$^{N\Delta38}$ and BipD$^{35\text{-}301}$ were overexpressed in *Escherichia coli* BL21 (DE3) (DNAY) grown in 1 liter of M9 minimal medium with ^{15}N ammonium chloride (Cambridge Isotope Laboratories, Inc.). The recombinant proteins were purified via Ni^{2+} affinity chromatography as described.[16] The His_6 tag was removed by TEV protease cleavage of purified His_6-GB1-(TEV)-SipD$^{N\Delta38}$ and His_6-(TEV)-BipD$^{35\text{-}301}$ as described.[17] The protein samples were concentrated using Amicon ultracel-3k (Millipore) and dialyzed into NMR buffer (10 mM sodium phosphate, pH 6.5, 10 mM NaCl). Protein concentration was estimated using absorbance at 280 nm.

Specific Amino Acid Labeling of Tip Protein

To facilitate the resonance assignment of SipD and BipD, 10 specific ^{15}N-labeled amino acids were used, such as ^{15}N-leucine, ^{15}N-valine, ^{15}N-isoleucine, ^{15}N-alanine, ^{15}N-phenoalanine, ^{15}N-glutamate, ^{15}N-aspartate, ^{15}N-tyrosine, ^{15}N-methionine and ^{15}N-lysine. The labeling of ^{15}N-glutamate and ^{15}N-aspartate, however, were unsuccessful due to scrambling of the ^{15}N label into other types of amino acids because both glutamate and aspartate are involved in the biosynthesis of other amino acids. Nevertheless, excluding Glycines and Prolines, 47% of total SipD residues

(excluding Prolines and Glycines) and 50% of total BipD residues were specifically labeled with ^{15}N-amino acids.

An *E.coli* starter culture was first grown in 1 liter LB media overnight, and was centrifuged and transferred into 500 ml M9 minimal medium containing various amounts of ^{15}N- and unlabeled amino acids in the media. After ~ 45 min equilibration, the bacteria were induced with 1 mM IPTG at A_{600} ~ 0.8 and cell growth was continued at 37 °C for 4 hours before harvest. The selective amino acid labeled recombinant proteins were purified same as unlabeled tip proteins as described above.

NMR Spectroscopy

NMR data were acquired at 30°C on a Bruker Avance 800 MHz equipped with cryogenic triple-resonance probes. Resonance assignments were obtained from 2D ^{1}H–^{15}N TROSY-HSQC[20] and the following 3D datasets: TROSY-HNCA,[21] TROSY-HNCACB[22] and TROSY-HN(CA)CO[22] using 0.46 mM perdeuterated ^{15}N,^{13}C-SipD$^{N\Delta38}$ and 0.4 mM perdeuterated ^{15}N,^{13}C-BipD$^{35\text{-}301}$ protein samples in NMR buffer (pH 6.5). NMR data were processed using NMRPipe[18] and analyzed using NMRView.[19]

NMR Chemical Shift Mapping

Recombinant ^{15}N-SipD$^{N\Delta38}$, ^{15}N-BipD$^{35\text{-}301}$, unlabeled PrgI$^{C\Delta5}$ and BsaL$^{C\Delta5}$ were purified as described and the concentration is determined by UV absorbance at

280 nm. NMR chemical shift mapping experiments were performed by acquiring 2D ^{1}H-^{15}N TROSY-HSQC at $PrgI^{C\Delta5}$:$SipD^{N\Delta38}$ molar ratios of 0 and 5. For BipD, NMR chemical shift mapping was performed at various $BsaL^{C\Delta5}$:$BipD^{35\text{-}301}$ molar ratios of 0, 0.27, 0.46, 0.92, 1.85 and 3.7. Both $PrgI^{C\Delta5}$ and $SipD^{N\Delta38}$ or $BsaL^{C\Delta5}$and $BipD^{35\text{-}301}$ were dialyzed into the same buffer prior to titration to assure that the chemical shift changes were only due to protein-protein interactions.

RESULTS

$BipD^{35\text{-}301}$ Residues Perturbed in BipD-BsaL Titration

NMR chemical shift mapping was used to identify the $BipD^{35\text{-}301}$ residues that bind $BsaL^{C\Delta5}$ because the interaction should result in chemical shift changes of $BipD^{35\text{-}301}$ residues. We titrated ^{15}N-labeled $BipD^{35\text{-}301}$ with unlabeled $BsaL^{C\Delta5}$ and monitored the changes in the chemical shifts of $BipD^{35\text{-}301}$ residues by acquiring a series of 2D TROSY-HSQC spectra with increasing amounts of $BsaL^{C\Delta5}$. Peak shifts were observed in the 2D HSQC of $BipD^{35\text{-}301}$ upon titration of $BsaL^{C\Delta5}$ (Fig. 5-5). The peaks shifted in one direction upon titration of $BsaL^{C\Delta5}$ indicating that BipD-BsaL interaction was in fast exchange NMR time scale, and weak interaction, in the order of 0.01 mM. Even when the molar ratio of $BsaL^{C\Delta5}$:$BipD^{35\text{-}301}$ reached around 4, the chemical shift changes were still relatively small, indicating weak interaction. However, the sequence specific assignment of $BipD^{35\text{-}301}$ is still ongoing and the residues that showed chemical shift changes have yet to be identified. A control

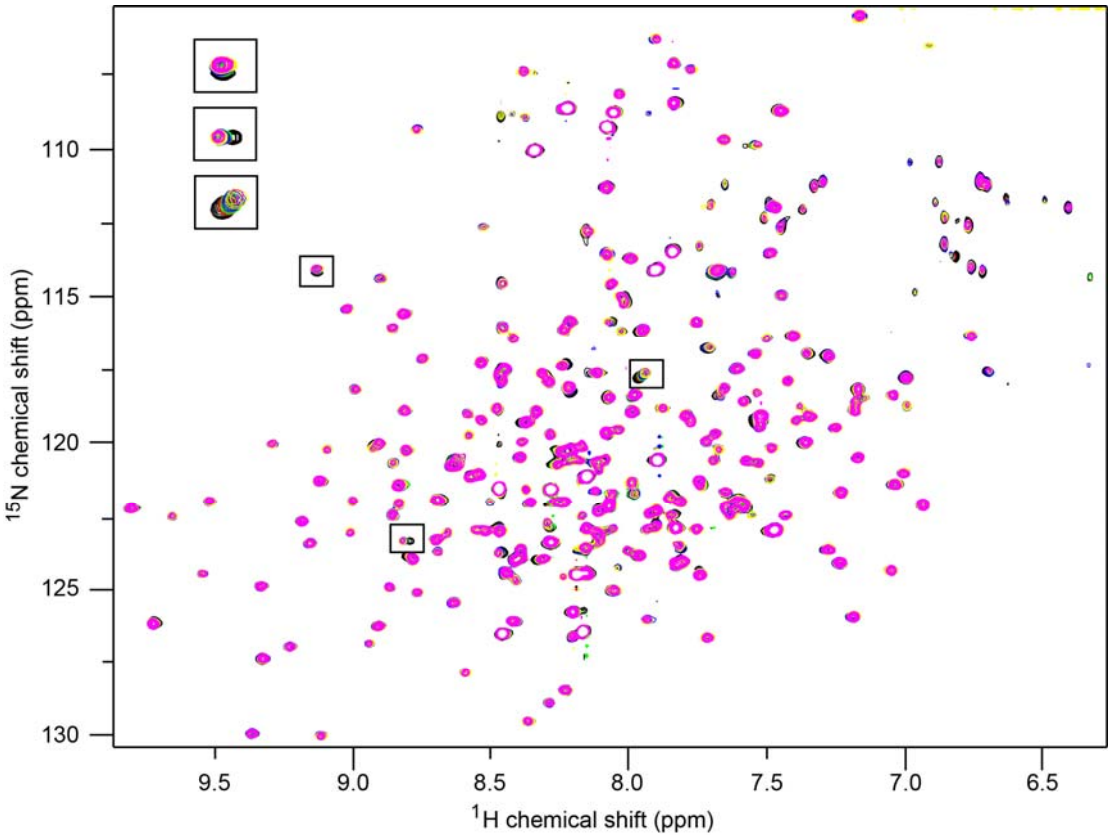

Figure 5-5. Titration of ^{15}N-BipD$^{35\text{-}301}$ with increasing amounts of unlabeled BsaL$^{C\Delta5}$. Overlay of six 2D HSQC spectra at various BsaL$^{C\Delta5}$:BipD$^{35\text{-}301}$ molar ratios (*black*, 0; *red*, 0.28; green, 0.45; blue, 1.0; yellow, 1.8 and magenta, 3.7). Clear peak shifts of some BipD$^{35\text{-}301}$ residues were observed upon addition of unlabeled BsaL$^{C\Delta5}$.

titration of mixing ^{15}N-BipD$^{35\text{-}301}$ with buffer was done at the dilution factor of 6 and no peak shifts were observed.

SipD-PrgI Interaction

We tried to use NMR chemical shift mapping to identify which SipD residues bind PrgI because our previous data showed that PrgI$^{C\Delta5}$ residues were perturbed when titrated with to SipD, despite weak interaction. The SipD-PrgI titration was done by adding unlabeled PrgI$^{C\Delta5}$ into ^{15}N-labeled SipD$^{N\Delta38}$. The changes in the chemical shifts of SipD$^{N\Delta38}$ peaks were monitored by acquiring a series of 2D TROSY-HSQC spectra with increasing amounts of PrgI. However, no obvious peak shifts were observed even at PrgI$^{C\Delta5}$:SipD$^{N\Delta38}$ molar ratio of 5 was reached (Fig. 5-6). A control titration of mixing ^{15}N-SipD$^{N\Delta38}$ with buffer was done at a dilution factor of 6 and no peak shift was observed either.

Sequence Specific Backbone Assignment of SipD$^{N\Delta38}$

We assigned the resonances of free SipD$^{N\Delta38}$ using a series of TROSY three-dimensional NMR experiments with the help of the specific amino acid labeling of 8 different amino acids (Fig. 5-7A). Assigning the NMR peaks of the 305-residue SipD$^{N\Delta38}$ was challenging because of the low sensitivity and poor resolution. SipD$^{N\Delta38}$ is predicted to be high α-helical, which is consistent with the presence of overcrowded peaks clustered withing a narrow proton chemical shift range. Nevertheless, using three-dimensional TROSY-HNCA, TROSY-HNCACB and

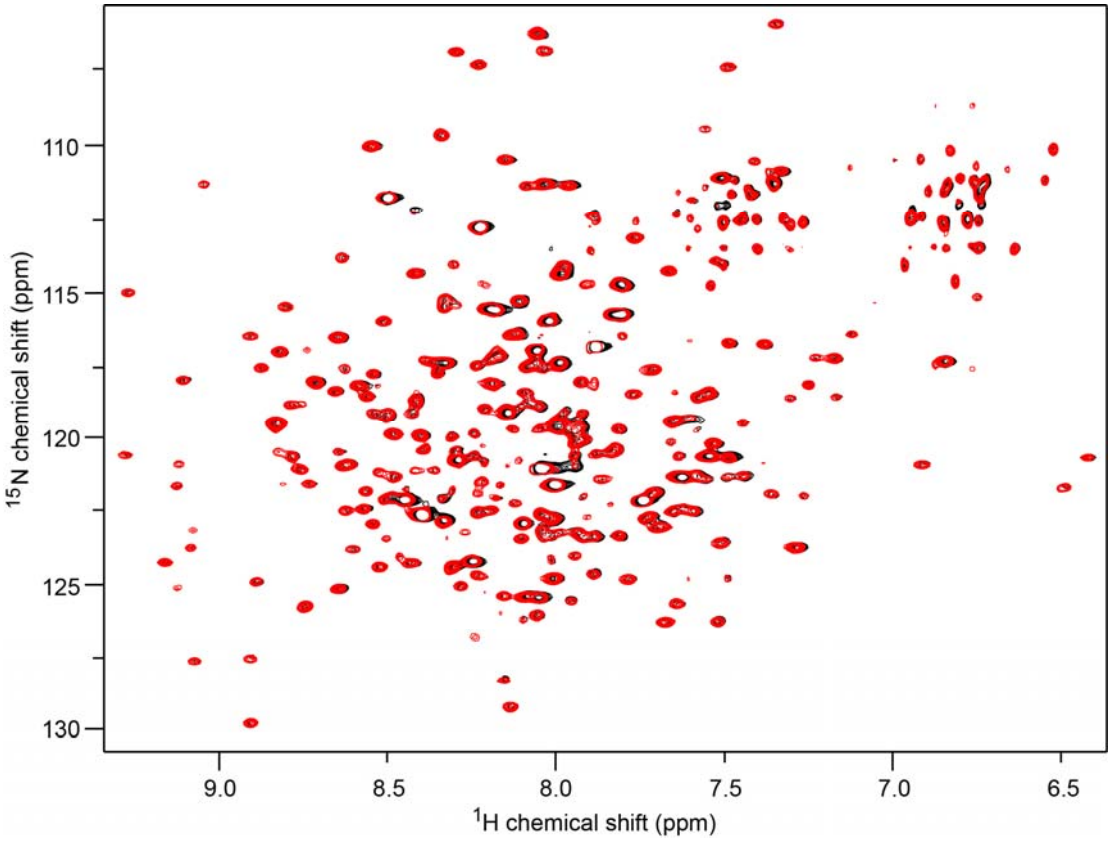

Figure 5-6. Titration of ^{15}N-SipD$^{N\Delta38}$ with increasing amounts of unlabeled PrgI$^{C\Delta5}$. Overlay of two 2D HSQC spectra at PrgI$^{C\Delta5}$:SipD$^{N\Delta38}$ molar ratios (*black*, 0 and *red*, 5). No obvious peak shifts of some SipD$^{N\Delta38}$ residues were observed upon addition of unlabeled BsaL$^{C\Delta5}$.

Figure 5-7A. Residues of SipD$^{N\Delta38}$ with backbone NMR assignments (shown in red, 216 out of 295 backbone amides were assigned). The structure is a predicted SipD model using the program I-TASSER.[23] Similar to IpaD structure, SipD is predicted to also adopte a dumbbell shape with a middle coiled-coil region that is flanked by two globular domains at two ends. The N-terminal 38 residues are in random coil and are omitted.

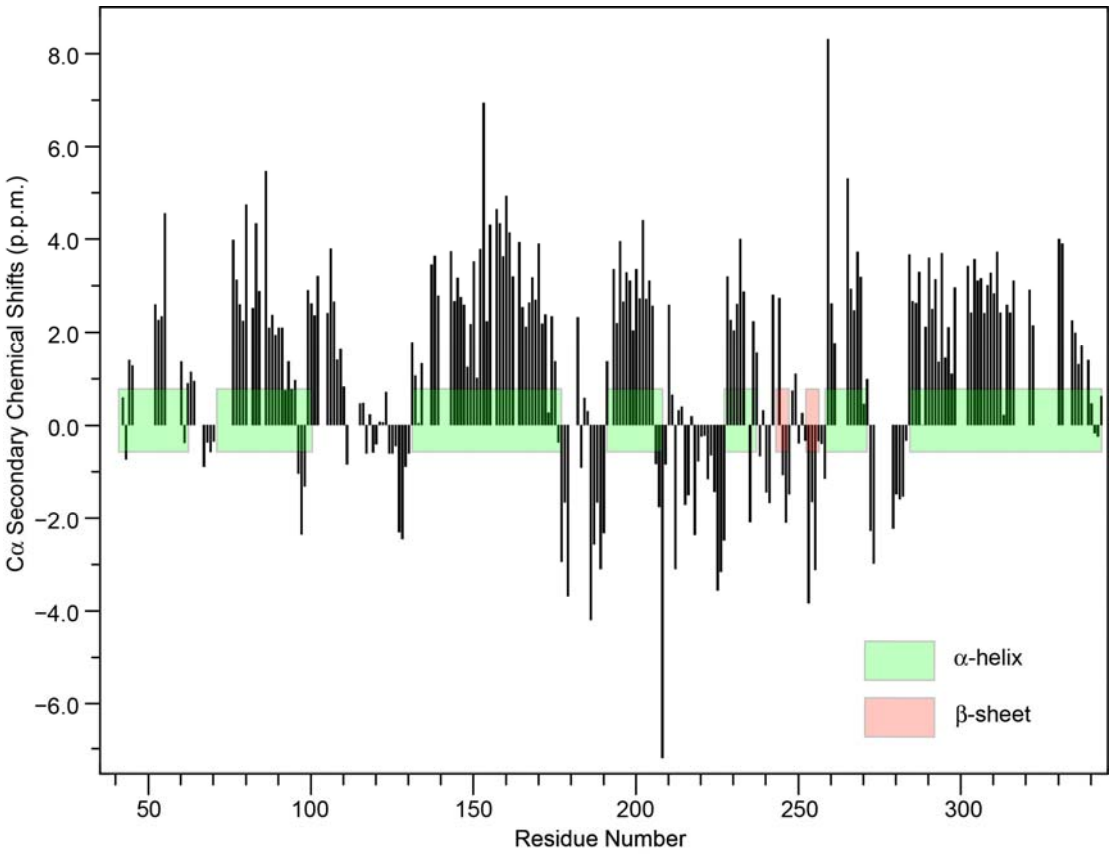

Figure 5-7B. Secondary chemical shifts obtained by substracting the random coil $^{13}C_{\alpha}$ chemical shift values from those observed in the protein indicate the regions of α-helices and β-strands. The regions shadowed in green indicate the boundary of α-helices and regions shadowed in red indicate the boundary of β-sheets in the predicted SipD model. Our secondary chemical shift index agrees with the predicted model and validates the model.

TROSY-HN(CA)CO, we assigned 216 of 295 (excluding the 10 prolines) backbone amides of $SipD^{N\Delta 38}$. The secondary C_α chemical shift index analysis indicates that SipD is likely to have a similar secondary structure as the predicted I-TASSER model (Fig. 5-7B).[23] We also collected the datasets for assigning $BipD^{35-301}$, but the assignment of $BipD^{35-301}$ are in slow progress due to the possible co-existence of both $BipD^{35-301}$ monomer and dimer in the solution.

DISCUSSION

The recruitment of tip protein on the top of bacterial needle is critical for bacterial invasion.[2] When the bacterial TTSA comes into direct contact with the cognate host cell,[24] the translocators are immediately secreted and inserted into the host cell membrane. The two translocators then form a translocon pore, thus opening a direct channel through which bacterial effectors can be injected into the host cell cytosol in a one-step process.[25] The Tip protein serves as a platform and is involved in multiple protein-protein interactions including needle-tip, tip-tip and tip-translocator interactions.

High solubility allows the biophysical and biochemical characterization of the tip proteins. Both SipD and BipD are soluble up to 0.7 mM without self-association or aggregating. Thus, it becames feasible to pursue studies of protein-protein interactions involving the tip protein by NMR spectroscopy. However, it is challenging for NMR studies of protein with more than 30 kDa, thus TROSY was

used to overcome this obstacle and TROSY versions of three dimensional experiments used in backbone assignments worked relatively well for the tip proteins.

The available atomic structures of LcrV[10], BipD[5; 11] and IpaD[5] show a high degree of α-helicity (Fig. 5-2). Similarly, SipD was also shown to be highly helical using CD spectroscopy.[9] The overlapping of peaks in a very narrow range of proton chemical shift in 2D TROSY-HSQC spectra is consistent with the predominantly α-helical structure of tip proteins (Fig. 5-5 and 5-6). The ambiguity of overlapped peaks of the helical region residues made sequence specific assignment of tip proteins relatively slow. Nevertheless, TROSY NMR methods combined with specific amino acid labeling resulted to assignment of over 70% of SipD residues. Most of the unassigned SipD residues belong to serine, glutamine or glutamate residues. The metabolic pathway of these three amino acids makes their specific labeling unlikely. Additional datasets including 3D and 4D experiments are needed in order to complete the assignment of SipD. The assignment of BipD is currently in progress.

NMR chemical shift mapping of SipD upon addition of PrgI did not yield any peak shifts. The reason for this was probably due to the weak interaction between PrgI and SipD. In order to characterize this weak interaction, Paramagnetic Relaxation Enhancement (PRE) will be used to determine the interaction between SipD and PrgI. PRE is more sensitive compared to chemical shift mapping. NMR chemical shift mapping of BipD upon titration with BsaL showed obvious peak shifts (Fig. 5-5). The progressively small shifts on each titration point indicate that the interaction between BipD and BsaL was weak and in fast exchange NMR time scale.

Sequence specific backbone assignment of BipD will reveal the residues that are perturbed by BsaL binding and this will also provide valuable information in understanding the tip-needle interaction in other Gram-negative bacterial systems. PrgI mutants will be tested using fluorescence polorization to correlate the NMR data. Further investigation will be done by *in vivo Salmonella* invasion assays and immunofluorescence staining of surface-exposed TTSA tip complex using SipD antibody. Western-blot analysis will be used to test the protein expression level and the secretion of proteins SipD, SipB and SipC into the culture supernatant.

It is proposed that the tip protein undergoes conformational changes at the N-terminal two helices, which swings away from the central coiled-coil and thereby allow binding to the needle protein.[15] It is possible that upon the shift of the N-terminal domain the tip protein will expose the surface which is important for the tip-tip interaction. The assignments of SipD provide a wealth of information not only for characterizing the tip-needle and tip-DOC interactions, but also the potential interaction of tip-tip and tip protein and translocator protein.

REFERENCES

1. Mueller, C. A., Broz, P., Muller, S. A., Ringler, P., Erne-Brand, F., Sorg, I., Kuhn, M., Engel, A. & Cornelis, G. R. (2005). The V-antigen of *Yersinia* forms a distinct structure at the tip of injectisome needles. *Science* **310**, 674-6.
2. Espina, M., Olive, A. J., Kenjale, R., Moore, D. S., Ausar, S. F., Kaminski, R. W., Oaks, E. V., Middaugh, C. R., Picking, W. D. & Picking, W. L. (2006). IpaD localizes to the tip of the type III secretion system needle of *Shigella flexneri*. *Infect Immun* **74**, 4391-400.
3. Picking, W. L., Nishioka, H., Hearn, P. D., Baxter, M. A., Harrington, A. T., Blocker, A. & Picking, W. D. (2005). IpaD of *Shigella flexneri* is

independently required for regulation of Ipa protein secretion and efficient insertion of IpaB and IpaC into host membranes. *Infect Immun* **73**, 1432-40.

4. Olive, A. J., Kenjale, R., Espina, M., Moore, D. S., Picking, W. L. & Picking, W. D. (2007). Bile salts stimulate recruitment of IpaB to the *Shigella flexneri* surface, where it colocalizes with IpaD at the tip of the type III secretion needle. *Infect Immun* **75**, 2626-9.
5. Johnson, S., Roversi, P., Espina, M., Olive, A., Deane, J. E., Birket, S., Field, T., Picking, W. D., Blocker, A. J., Galyov, E. E., Picking, W. L. & Lea, S. M. (2007). Self-chaperoning of the type III secretion system needle tip proteins IpaD and BipD. *J Biol Chem* **282**, 4035-44.
6. Ide, T., Laarmann, S., Greune, L., Schillers, H., Oberleithner, H. & Schmidt, M. A. (2001). Characterization of translocation pores inserted into plasma membranes by type III-secreted Esp proteins of enteropathogenic *Escherichia coli*. *Cell Microbiol* **3**, 669-79.
7. Wang, Y., Zhang, L., Picking, W. L., Picking, W. D. & Guzman, R. N. D. (2008). Structural dissection of the extracellular moieties of the type III secretion apparatus. *Mol. BioSyst.* **4**, 1176-80.
8. Kaniga, K., Trollinger, D. & Galan, J. E. (1995). Identification of two targets of the type III protein secretion system encoded by the inv and spa loci of *Salmonella typhimurium* that have homology to the *Shigella* IpaD and IpaA proteins. *J Bacteriol* **177**, 7078-85.
9. Espina, M., Ausar, S. F., Middaugh, C. R., Baxter, M. A., Picking, W. D. & Picking, W. L. (2007). Conformational stability and differential structural analysis of LcrV, PcrV, BipD, and SipD from type III secretion systems. *Protein Sci* **16**, 704-14.
10. Derewenda, U., Mateja, A., Devedjiev, Y., Routzahn, K. M., Evdokimov, A. G., Derewenda, Z. S. & Waugh, D. S. (2004). The structure of *Yersinia pestis* V-antigen, an essential virulence factor and mediator of immunity against plague. *Structure* **12**, 301-6.
11. Erskine, P. T., Knight, M. J., Ruaux, A., Mikolajek, H., Wong Fat Sang, N., Withers, J., Gill, R., Wood, S. P., Wood, M., Fox, G. C. & Cooper, J. B. (2006). High resolution structure of BipD: an invasion protein associated with the type III secretion system of *Burkholderia pseudomallei*. *J Mol Biol* **363**, 125-36.
12. Broz, P., Mueller, C. A., Muller, S. A., Philippsen, A., Sorg, I., Engel, A. & Cornelis, G. R. (2007). Function and molecular architecture of the *Yersinia* injectisome tip complex. *Mol Microbiol* **65**, 1311-20.
13. Deane, J. E., Roversi, P., Cordes, F. S., Johnson, S., Kenjale, R., Daniell, S., Booy, F., Picking, W. D., Picking, W. L., Blocker, A. J. & Lea, S. M. (2006). Molecular model of a type III secretion system needle: Implications for host-cell sensing. *Proc Natl Acad Sci U S A* **103**, 12529-33.
14. Caroline, G., Eric, F., Bohn, Y. S., Sylvie, E. & Attree, I. (2008). Oligomerization of PcrV and LcrV, protective antigens of *Pseudomonas aeruginosa* and *Yersinia pestis*. *J Biol Chem* **283**, 23940-9.

15. Zhang, L., Wang, Y., Olive, A. J., Smith, N. D., Picking, W. D., De Guzman, R. N. & Picking, W. L. (2007). Identification of the MxiH needle protein residues responsible for anchoring invasion plasmid antigen D to the type III secretion needle tip. *J Biol Chem* **282**, 32144-51.
16. Wang, Y., Ouellette, A. N., Egan, C. W., Rathinavelan, T., Im, W. & De Guzman, R. N. (2007). Differences in the electrostatic surfaces of the type III secretion needle proteins PrgI, BsaL, and MxiH. *J Mol Biol* **371**, 1304-14.
17. Wang, Y., Boudreaux, D. M., Estrada, D. F., Egan, C. W., St Jeor, S. C. & De Guzman, R. N. (2008). NMR structure of the N-terminal coiled coil domain of the Andes hantavirus nucleocapsid protein. *J Biol Chem* **283**, 28297-304.
18. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-93.
19. Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* **278**, 313-52.
20. Czisch, M. & Boelens, R. (1998). Sensitivity Enhancement in the TROSY Experiment *J. Magn. Reson.* **134**, 158-160.
21. Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wuthrich, K. (1998). TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci U S A* **95**, 13585-90.
22. Salzmann, M., Wider, G., Pervushin, K., Senn, H. & Wuethrich, K. (1999). TROSY-type Triple-Resonance Experiments for Sequential NMR Assignments of Large Proteins. *J. Am. Chem. Soc.* **121**, 844-848.
23. Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40.
24. Rosqvist, R., Magnusson, K. E. & Wolf-Watz, H. (1994). Target cell contact triggers expression and polarized transfer of *Yersinia* YopE cytotoxin into mammalian cells. *EMBO J* **13**, 964-72.
25. Bahrani, F. K., Sansonetti, P. J. & Parsot, C. (1997). Secretion of Ipa proteins by *Shigella* flexneri: inducer molecules and kinetics of activation. *Infect Immun* **65**, 4005-10.
26. Stensrud, K. F., Adam, P. R., La Mar, C. D., Olive, A. J., Lushington, G. H., Sudharsan, R., Shelton, N. L., Givens, R. S., Picking, W. L. & Picking, W. D. (2008). Deoxycholate interacts with IpaD of *Shigella flexneri* in inducing the recruitment of IpaB to the type III secretion apparatus needle tip. *J Biol Chem* **283**, 18646-54.

Chapter 6: Characterization of SipD-Deoxycholate Interaction

INTRODUCTION

The maturation of TTS (type III secretion) was suggested to be a step-wise process.[1] In *Shigella*, IpaD localizes to the tip of the *Shigella* needle and controls the secretion of translocators IpaB and IpaC.[2] IpaD was further shown to respond to the DOC (deoxycholate) and other bile salts as well, to promote the stable recruitment of translocator protein IpaB to the needle tip without inducing a rapid burst of type III secretion.[1] Only IpaB, not IpaC, was shown to colocalize with IpaD in the presence of DOC, suggesting that further recruitment of IpaC on the tip of bacterial needle requires further activation signal the type III secretion.[1] This is consistent with the observation that the type III secretion is controlled by both IpaD and IpaB.[1; 3] DOC bound to IpaD was shown to induce the bacteria into a secretion-competent state and may represent the first step of the step-wise maturation of the type III secretion..[1] Membrane protein IpaB was shown to bind with cholesterol[4], and combining the fact that IpaB interacts with the host cell membrane lipid rafts enriched with cholesterols and sphingolipids at which the invasion occurs[1; 5; 6], it is likely the IpaB will play a role to further activate the type III secretion.

In *Shigella*, it is shown that bile salts induce the protein secretion and enhance the bacterial invasion.[7] Opposite effect was observed that the *Salmonella* SPI-1 invasion gene transcriptions were suppressed in the presence of Bile, and the *Salmonella typhimurium* invasion was greatly repressed.[8; 9] Although information on

whether bile salt plays a similar role to activate the type III secretion in *Salmonella* remains unknown, it is suggested that the SPI-1 gene transcription would be derepressed and the type III secretion will be initiated in low bile concentration upon the bacterial penetration of the mucous layer.[8]

DOC is shown to bind to IpaD and SipD. The fluorescence polarization data showed DOC binds to IpaD with around 10 μM disassociation constant, with a lower binding affinity with SipD. Using computer docking methods, DOC is suggested to bind IpaD along one side of the central coiled-coil that is close to the N-terminal two helix boundle (Fig. 6-S1).[10] This model is consistent with the FRET (*Förster Resonance Energy Transfer*) data which suggests DOC binds close to the midpoint of the protein.[10] The positioning suggests IpaB has directly contact with IpaD, however, no direct evidence has been shown that IpaB binds IpaD. It was suggested that DOC is bound by IpaD to induce a conformational change in IpaD, and this allows the secretion of IpaB, which is then bound to IpaD at the needle tip.[1] The close evolutionary similarity between *Salmonella* and *Shigella* suggests a similar activation mechanism upon DOC binding. Membrane protein IpaB with a molecular weight of 62 kDa makes the characterization of IpaD-IpaB interaction challenging. Currently available data suggest IpaB binds IpaD first, and then IpaC binds IpaB to form the translocon pore on the host cell surface.

In this chapter, we characterized the SipD-DOC interaction using NMR chemical shift mapping. The DOC binding site of SipD was mapped to the bottom of the central coiled-coil and the long loop connected to the N-terminal domain, on the

opposite side of coiled-coil where locates the N-terminal two helices. This binding site, however, is different from the binding site obtained by computer simulation for IpaD-DOC interaction. This may suggest a different activation mechanism of type III secretion between *Salmonella* and *Shigella*, or probably reflects DOC binding to a weaker affinity site in SipD, and a strong affinity site in IpaD.[10] Our data provide preliminary information for understanding the step-wise maturation of type III secretion secretion in *Slamonella*.

MATERIALS AND METHODS

Protein Expression and Purification

Isotopically (^{15}N) labeled $SipD^{N\Delta 38}$ and $BipD^{35\text{-}301}$ were overexpressed in *Escherichia coli* BL21(DE3)(DNAY) grown in 1 liter of M9 minimal medium with ^{15}N ammonium chloride (Cambridge Isotope Laboratories, Inc.). The recombinant proteins were purified via Ni^{2+} affinity chromatography as described.[11] The His_6 tag was removed by TEV protease cleavage of purified His_6-GB1-(TEV)-$SipD^{N\Delta 38}$ and His_6-(TEV)-$BipD^{35\text{-}301}$ as described.[12] The protein samples were concentrated using Amicon ultracel-3k (Millipore) and dialyzed into NMR buffer (10 mM sodium phosphate, pH 6.5, 10 mM NaCl). Protein concentration was estimated using absorbance at 280 nm.

NMR Spectroscopy

NMR data were acquired at 30°C on a Bruker Avance 800 MHz equipped with cryogenic triple-resonance probes. NMR data were obtained using 2D $^{1}H–^{15}N$ TROSY-HSQC experiment[13]. NMR data were processed using NMRPipe[14] and analyzed using NMRView.[15]

NMR Chemical Shift Mapping

Recombinant ^{15}N-SipD$^{N\Delta 38}$ and ^{15}N-BipD$^{35-301}$were purified as described and the concentration is determined by UV absorbance at 280nm. The chemical shift mapping of DOC bound to tip proteins was done by dialyzing the 0.61 mM ^{15}N-SipD$^{N\Delta 38}$ and 1.1 mM ^{15}N-BipD$^{35-301}$ into NMR buffer with 0M, 0.7 mM, 1.4 mM and 2.1 mM DOC, respectively. Control titrations were done by titrating SipD$^{N\Delta 38}$ and BipD$^{35-301}$ with NMR buffer at the dilution factor of 6.

RESULTS

SipD$^{N\Delta 38}$ Residues Perturbed in the SipD-DOC Titration

It was shown that the bile salt deoxycholate (DOC) interacts with IpaD and SipD but not to BipD.[10] We used NMR chemical shift mapping to identify the SipD residues that are perturbed upon titration with DOC. We dialyzed ^{15}N-labeled SipD$^{N\Delta 38}$ into NMR buffer with varying concentrations of DOC and monitored the changes in the chemical shifts of SipD$^{N\Delta 38}$ by acquiring a series of 2D TROSY-HSQC at DOC:SipD$^{N\Delta 38}$ molar ratios of 0, 1, 2 and 3. Peak shifts were observed in the 2D HSQC of SipD$^{N\Delta 38}$ upon addition of DOC (Fig. 6-1). Further increase in the amount

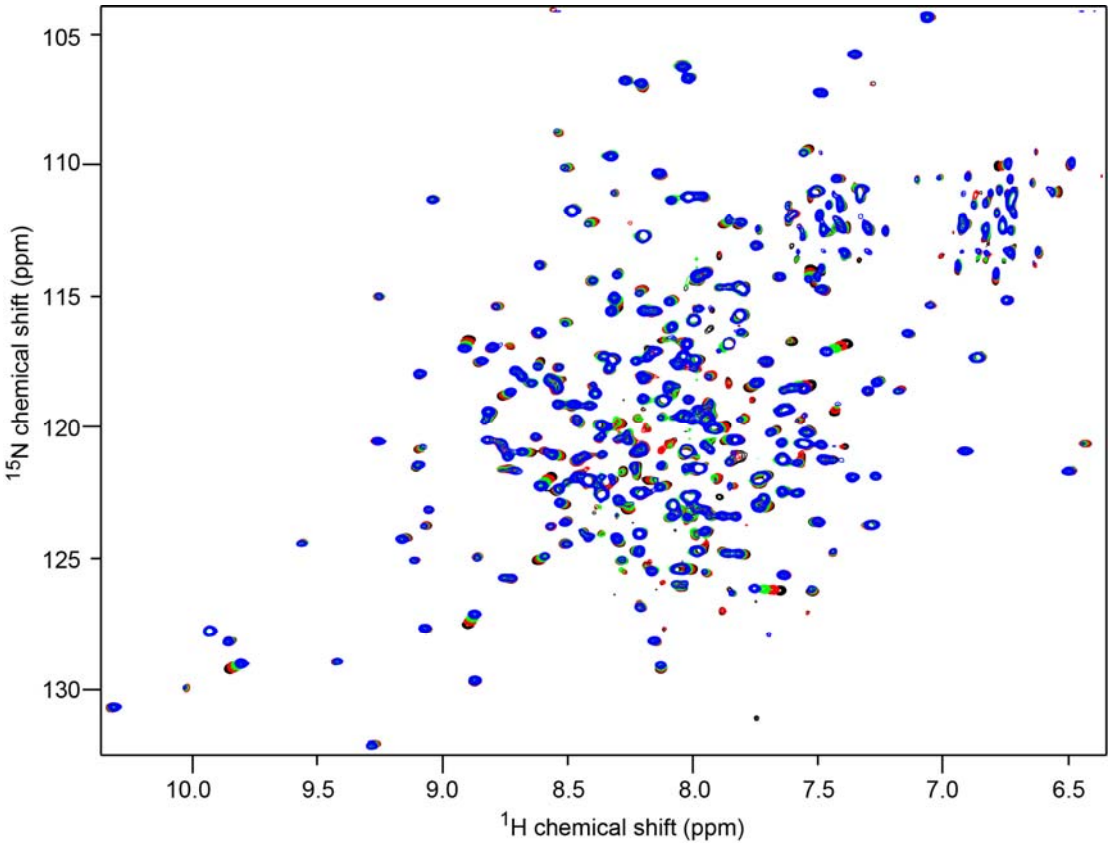

Figure 6-1. Titration of ^{15}N-SipD$^{N\Delta38}$ with increasing amounts of deoxycholate (DOC). Overlay of four 2D HSQC spectra at various DOC:SipD$^{N\Delta38}$ molar ratios (*black*, 0; *red*, 1; *green*, 2 and *blue*, 3). Clear shifts of some SipD$^{N\Delta38}$ peaks were observed with increasing amount of DOC.

of DOC changed the peak positions even further in one direction, indicating that binding was in fast exchange NMR time scale.

Using the sequence specific assignment of $SipD^{N\Delta 38}$ described in Chapter 6, we were able to identify most the residues that were perturbed upon titration with DOC. Residues that showed chemical shift changes are Ser96, Phe109, Ser110, Ala111, Leu116, Phe117, Ala119, Gln124, Asp131, Glu133, Tyr149, Gln165, Leu179, Val187, Lys188, Val256, Gln306, Thr307 and Leu322. Along with residues that showed relatively small chemical shift changes: Ser98, Thr127, Ile142, Ala144, Ser148, Asn155, Ala158, Glu237, Leu238, Met282, Gln291 and Thr303. A binding site is located on the bottom of the central coiled-coil and the extended loop between the coiled-coil and the N-terminal domain, on the reverse side of the central coiled-coil where the N-terminal two helices locate (Fig. 6-2). Although the atomic structure of SipD is unavailable, most of the residues on the DOC binding are clustered on one site, which suggests a clear binding site of SipD for DOC. As control, similar DOC titration was performed for $BipD^{35\text{-}301}$, however, no obvious peak shifts were observed, indicating that there was no interaction between DOC and BipD, as initially reported.[10]

DISCUSSION

The tip protein was shown to assemble into a ring-like structure at the tip of the needle[16] and recruitment of the tip protein is critical for the bacterial invasion.[2] DOC was shown to bind *Shigella* tip protein IpaD, and IpaD further recruits the

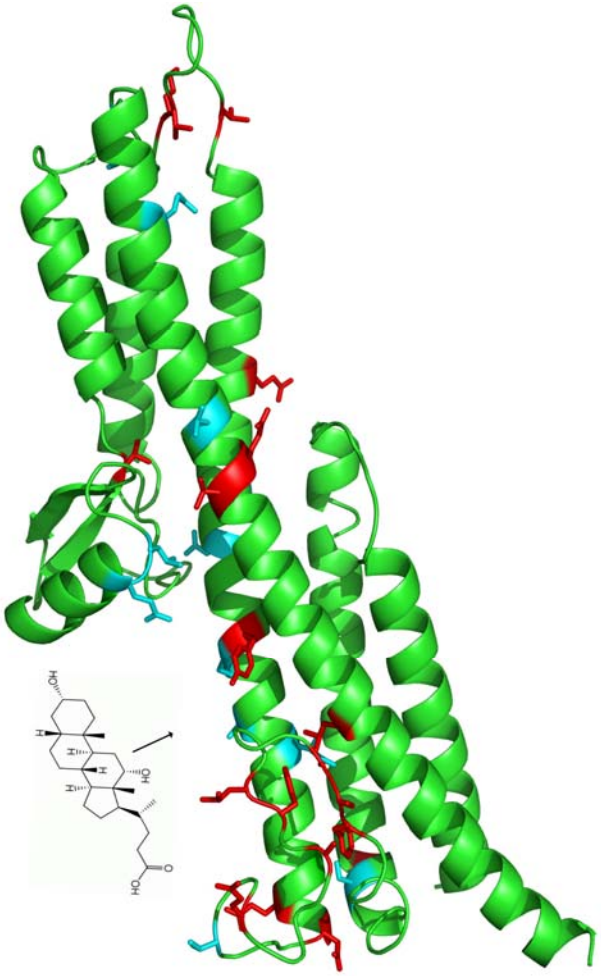

Figure 6-2. Perturbed residues of SipD$^{N\Delta38}$ (shown as sticks) upon titration of DOC, using the predicted I-TASSER model.[17] Residues labeled in red showed significant chemical shift changes and residues labeled in cyan showed relatively smaller chemical shift changes. A binding site for DOC clustered at the bottom of the coiled-coil, on the reverse side of coiled-coil which has the N-terminal two helices.

translocator protein IpaB at the tip of the needle.[1] This is the initial sensing of the environment signal for the activation of the type III secretion.[1] It was suggested the IpaB acted as a sensor for further activation signal and was likely to bind to the host membrane and activate the type III secretion upon host cell contact.[1; 18] However, further information on how IpaC is recruited and inserted into the host cell membrane still remains unclear.

Although bile was shown to repress the type III secretion by suppressing the SPI-1 gene transcription, a similar activation mechanism was adopted by *Salmonella* upon DOC binding.[8; 9] *In vitro* data showed that both IpaD and SipD were able to bind DOC.[10] We used NMR chemical shift mapping to characterize the SipD-DOC binding, and the SipD-DOC interaction was shown to be weak. Using the known assignments of SipD that have been completed to date, we were able to identify residues that are perturbed upon the titration of DOC. These residues were clustered along one side of the central coiled-coil (Fig. 6-2), which is suggested to be the binding pocket for DOC. Due to incomplete assignment, not all of the residues close to the binding site were shown to have chemical shift changes. An earlier model proposed the binding site of IpaD with DOC is located at the reverse side of the coiled-coil and between the central coiled-coil and the N-terminal two helices.[10] It was also found that DOC bound more tightly to IpaD compared to SipD, thus the binding site ascribed to the SipD-DOC interaction might represent the weaker affinity site of DOC.[10] It will be interesting to determine the binding site of IpaD with DOC using chemical shift mapping experiment. There are some other perturbed residues

(Leu179, Lys188 and Met 282) located at the loop regions between distal domain and the central coiled-coil. The perturbation of these SipD residues is probably due to the induced conformational changes instead of direct contact with DOC. This conformational change might facilitate the subsequent sensing of the the type III secretion activation signal and a possible scenario could be that the distal domain swings away from the central coiled-coil which then facilitates the recruitment of translocator protein IpaB at the tip of the type III secretion needle. Domain dynamics of SipD upon binding to DOC will be characterized by NMR. The possible protein-protein interactions between the tip protein and the translocator protein will be studied as well. Information of loop dynamics and the tip-translocator interaction will be useful for understanding the further recruitment of the translocator and the activation mechanism of type III secretion.

REFERENCES

1. Olive, A. J., Kenjale, R., Espina, M., Moore, D. S., Picking, W. L. & Picking, W. D. (2007). Bile salts stimulate recruitment of IpaB to the *Shigella flexneri* surface, where it colocalizes with IpaD at the tip of the type III secretion needle. *Infect Immun* **75**, 2626-9.
2. Espina, M., Olive, A. J., Kenjale, R., Moore, D. S., Ausar, S. F., Kaminski, R. W., Oaks, E. V., Middaugh, C. R., Picking, W. D. & Picking, W. L. (2006). IpaD localizes to the tip of the type III secretion system needle of *Shigella flexneri*. *Infect Immun* **74**, 4391-400.
3. Menard, R., Sansonetti, P. & Parsot, C. (1994). The secretion of the *Shigella flexneri* Ipa invasins is activated by epithelial cells and controlled by IpaB and IpaD. *EMBO J* **13**, 5293-302.
4. Hayward, R. D., Cain, R. J., McGhie, E. J., Phillips, N., Garner, M. J. & Koronakis, V. (2005). Cholesterol binding by the bacterial type III translocon is essential for virulence effector delivery into mammalian cells. *Mol Microbiol* **56**, 590-603.

5. Lafont, F., Tran Van Nhieu, G., Hanada, K., Sansonetti, P. & van der Goot, F. G. (2002). Initial steps of *Shigella* infection depend on the cholesterol/sphingolipid raft-mediated CD44-IpaB interaction. *EMBO J* **21**, 4449-57.
6. van der Goot, F. G., Tran van Nhieu, G., Allaoui, A., Sansonetti, P. & Lafont, F. (2004). Rafts can trigger contact-mediated secretion of bacterial effectors via a lipid-based mechanism. *J Biol Chem* **279**, 47792-8.
7. Pope, L. M., Reed, K. E. & Payne, S. M. (1995). Increased protein secretion and adherence to HeLa cells by *Shigella* spp. following growth in the presence of bile salts. *Infect Immun* **63**, 3642-8.
8. Prouty, A. M. & Gunn, J. S. (2000). *Salmonella enterica* serovar typhimurium invasion is repressed in the presence of bile. *Infect Immun* **68**, 6763-9.
9. Prouty, A. M., Brodsky, I. E., Falkow, S. & Gunn, J. S. (2004). Bile-salt-mediated induction of antimicrobial and bile resistance in *Salmonella typhimurium*. *Microbiology* **150**, 775-83.
10. Stensrud, K. F., Adam, P. R., La Mar, C. D., Olive, A. J., Lushington, G. H., Sudharsan, R., Shelton, N. L., Givens, R. S., Picking, W. L. & Picking, W. D. (2008). Deoxycholate interacts with IpaD of *Shigella flexneri* in inducing the recruitment of IpaB to the type III secretion apparatus needle tip. *J Biol Chem* **283**, 18646-54.
11. Wang, Y., Ouellette, A. N., Egan, C. W., Rathinavelan, T., Im, W. & De Guzman, R. N. (2007). Differences in the electrostatic surfaces of the type III secretion needle proteins PrgI, BsaL, and MxiH. *J Mol Biol* **371**, 1304-14.
12. Wang, Y., Boudreaux, D. M., Estrada, D. F., Egan, C. W., St Jeor, S. C. & De Guzman, R. N. (2008). NMR structure of the N-terminal coiled coil domain of the Andes hantavirus nucleocapsid protein. *J Biol Chem* **283**, 28297-304.
13. Czisch, M. & Boelens, R. (1998). Sensitivity Enhancement in the TROSY Experiment *J. Magn. Reson.* **134**, 158-160.
14. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-93.
15. Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* **278**, 313-52.
16. Mueller, C. A., Broz, P., Muller, S. A., Ringler, P., Erne-Brand, F., Sorg, I., Kuhn, M., Engel, A. & Cornelis, G. R. (2005). The V-antigen of *Yersinia* forms a distinct structure at the tip of injectisome needles. *Science* **310**, 674-6.
17. Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40.
18. Rosqvist, R., Magnusson, K. E. & Wolf-Watz, H. (1994). Target cell contact triggers expression and polarized transfer of *Yersinia* YopE cytotoxin into mammalian cells. *EMBO J* **13**, 964-72.

SUPPLEMENTAL FIGURES

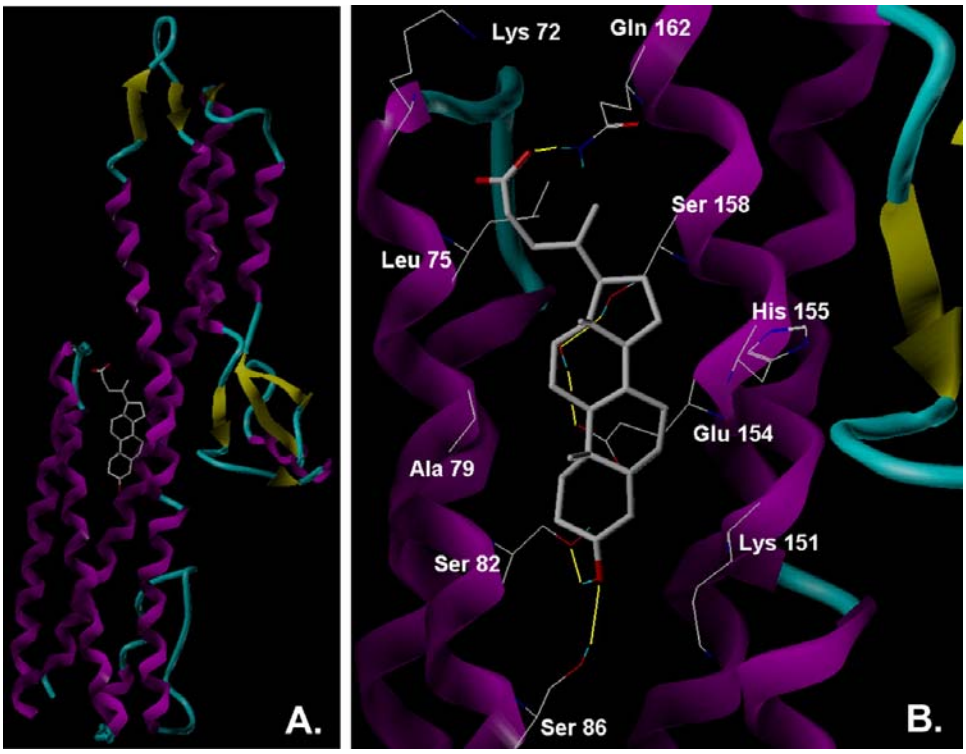

Figure 5-S1. Simulated docking of deoxycholate onto the crystal structure of IpaD is shown. (**A**) deoxycholate docked onto the ribbon structure of IpaD. (**B**) enlarged version of the binding pocket showing key contacts.[10] (Fig 5-S1 is from Ref 10.)

Chapter 7: Summary of TTSA Project

The type III secretion system (TTSS) is commonly utilized by many Gram-negative bacteria to infect host cells.[1; 2] Among these bacteria are human pathogens, which can cause a variety of infectious diseases.[3] The bacterial pathogenicity is mediated by the disruption of normal cellular function by effector proteins, which is injected from the bacterium into the host cell through a TTSS mediated process.[1; 3] Although effector proteins diverge from pathogen to pathogen[4; 5], the type III secretion apparatus (TTSA), which is the multiple-protein assembly for this protein secretion process, is structurally and functionally conserved.[6] The TTSA is composed of basal body which spans both bacterial inner membrane and outer membrane, and the external needle topped with a tip complex that controls the secretion and delivery of the translocators and effector proteins. The structure and function studies of TTSA, especially the assembly of needle and tip complex, are necessary in understanding the type III secretion mechanism.[6; 7] Despite an increasing body of work on the high resolution structures of individual proteins[8; 9; 10; 11; 12; 13; 14; 15; 16], gaps in knowledge remain on how proteins interact and assemble.[7; 8] We used NMR spectroscopy to study the protein structures, protein dynamics and protein-protein interactions involved in the assembly of TTSA needle and tip complex.

The NMR structures of both BsaL$^{C\Delta5}$ and PrgI$^{C\Delta5}$ showed a common central two helix bundle flanked by N- and C-terminal flexible tails.[12; 13] The flexible tails adopt partial α-helical character with intermediate rigidity in both BsaL$^{C\Delta5}$ and

PrgI$^{C\Delta5}$, suggesting that upon protein-protein interaction those flexible termini will experience structural changes and form regular α-helix.[12; 13] The electrostatic surface potential analysis of PrgI$^{C\Delta5}$, BsaL$^{C\Delta5}$ and MxiH$^{C\Delta5}$ showed radical differences in the negatively charged surfaces.[12] Despite being highly acidic proteins, these proteins contain positively charged surfaces, suggesting electrostatic contacts are important in needle assembly.[12] It will be interesting to determine the specific ionic strength and salt compositions that could disrupt needle assembly; and this exploration will not be limited to small molecules but also small peptides. Modeling of the needle suggests that the *Salmonella* needle might assemble differently compared to the *Shigella* needle.[12] The exact topology of *Salmonella* needle will require the high resolution structure determination of the intact needle.

We also used the method of NMR chemical shift mapping to characterize the PrgI monomer-monomer interaction. Residues in various sites of PrgI were perturbed in the titration experiments, suggesting multiple contacts are involved in the PrgI-PrgI interaction. This result is consistent with the proposed model by Deane *et al*. that a needle monomer is surrounded by 7 needle monomers, indicating inter-molecule interactions. Many of the PrgI residues perturbed upon PrgI-PrgI titration are charged residues, suggesting that electrostatic interaction could be important in PrgI-PrgI interaction. Some hydrophobic residues are also perturbed in the PrgI-PrgI titration, suggesting hydrophobic interaction might also contribute to needle assembly. Additionally, our NMR titration data showed a weak binding interaction between PrgI monomers, in the order of K_d ~ 0.7 mM. This weak binding is not surprising

considering that needle monomers are packed against other needle monomers in a super-helical fashion; and that high affinity interaction could be achieved by a high local needle monomer concentration. By using truncation mutant ($PrgI^{C\Delta5}$) or GB1-fusion to obtain monomeric proteins for solution studies, our data, however do not tell the native dissociation constant among the intact needle monomers, which might be much higher than the affinity suggested by our data. Further characterizations including fine distance restraints between needle monomers using NOE (nuclear Overhauser effect), RDC (residual dipolar coupling) and PRE (paramagnetic relaxation enchancement) restraints are needed. High resolution structural techniques like cryo-EM and solid-state NMR will also provide unique information on intact needles.

The PrgI-SipD NMR titration data showed distinct sites in PrgI that are involved in the PrgI-SipD interaction, compared to the PrgI-PrgI interaction. Residues on the PrgI C-terminal tail were shown to be involved the PrgI-SipD interaction. Residues at the N-terminus were also perturbed; and this is likely due to the allosteric conformational change upon the PrgI-SipD interaction. Although the chemical shift mapping of SipD-PrgI titration did not show any obvious peak shifts, we were able to show tip-needle interaction using *Burkholderia* homologous proteins BipD and BsaL. After the completion of the backbone resonance assignment of BipD, the binding sites of BipD will be revealed upon the binding to BsaL; and this information will provide general knowledge on the bacterial tip-needle interaction. The weak interactions between bacterial needle proteins and tip proteins will be studied using

paramagnetic relaxation enhancement (PRE), an NMR technique that relies on the use of spin labels, which can identify weak protein-protein interactions. Although this weak needle-tip interaction might not represent the native interaction between the assembled needle and tip complex, further studies on how the tip proteins assemble into the tip complex will be characterized as well.

Recently, binding of deoxycholate (DOC) to IpaD is suggested to be an early step during the activation of the type III secretion in *Shigella flexneri*.[17] The availability of tip proteins that gave excellent NMR spectra indicated that the tip protein-DOC interactions can be studied by NMR. The titration of SipD with DOC showed perturbed SipD residues that were clustered into one site located at the bottom of the central coiled-coil and the long loop connected to the N-terminal domain, on the opposite side of coiled-coil where locates the N-terminal two helices. There are some other residues located at the loop regions, suggesting possible conformational changes of both the N-terminal two-helix domain and the distal domain. Our result is different with the proposed computer model of DOC binding to IpaD[18], although this difference might not exclude the possibility that same activation mechanism upon DOC binding is used for both *Shigella* and *Salmonella*. It was also found that DOC bound more tightly to IpaD compared to SipD[18], thus the binding site ascribed to the SipD-DOC interaction might represent the weaker affinity site of DOC. The full NMR resonance assignments of SipD will provide a wealth of information not only for characterizing the tip-needle and tip-DOC interactions, but also open the possibility of studies addressing tip-tip and tip-translocator interactions.

In summary, the data in this dissertation show primary information for understanding the needle assembly by solving the structure of the needle protein, PrgI.[12] Both N- and C-termini of $PrgI^{C\Delta5}$ and $BsaL^{C\Delta5}$ adopt semi-rigid partial α-helical character.[12; 13] We suggested that soluble needle proteins might possess different structure dynamics compared to the assembled needle monomers. Our data also showed weak interactions were involved in the needle monomer-monomer interaction and needle-tip interaction. Most likely, this weak interaction arose from the changes made in the needle protein so it can be solubilized. Distinct binding sites of PrgI were involved in the PrgI-PrgI and the PrgI-SipD interactions. These weak interactions provide a starting point for the design the anti-bacterial agents to disrupt these protein-protein interactions involved in the assembly of type III secretion apparatus. When DOC binds to the tip protein, the tip protein is suggested to experience conformational changes, which will allow the further recruitment of the translocator protein. Studies on how the interaction between the tip protein and the translocator protein will provide additional structural information for further understanding the various stages in the maturation of the type III secretion apparatus.

REFERENCES

1. Hueck, C. J. (1998). Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev* **62**, 379-433.
2. He, S. Y., Nomura, K. & Whittam, T. S. (2004). Type III protein secretion mechanism in mammalian and plant pathogens. *Biochim Biophys Acta* **1694**, 181-206.
3. Coburn, B., Sekirov, I. & Finlay, B. B. (2007). Type III secretion systems and disease. *Clin Microbiol Rev* **20**, 535-49.

4. Cheng, L. W. & Schneewind, O. (2000). Type III machines of Gram-negative bacteria: delivering the goods. *Trends Microbiol* **8**, 214-20.
5. Galan, J. E. (2001). *Salmonella* interactions with host cells: type III secretion at work. *Annu Rev Cell Dev Biol* **17**, 53-86.
6. Blocker, A., Komoriya, K. & Aizawa, S. (2003). Type III secretion systems and bacterial flagella: insights into their function from structural similarities. *Proc Natl Acad Sci U S A* **100**, 3027-30.
7. Wang, Y., Boudreaux, D. M., Estrada, D. F., Egan, C. W., St Jeor, S. C. & De Guzman, R. N. (2008). NMR structure of the N-terminal coiled coil domain of the Andes hantavirus nucleocapsid protein. *J Biol Chem* **283**, 28297-304.
8. Deane, J. E., Roversi, P., Cordes, F. S., Johnson, S., Kenjale, R., Daniell, S., Booy, F., Picking, W. D., Picking, W. L., Blocker, A. J. & Lea, S. M. (2006). Molecular model of a type III secretion system needle: Implications for host-cell sensing. *Proc Natl Acad Sci U S A* **103**, 12529-33.
9. Johnson, S., Roversi, P., Espina, M., Olive, A., Deane, J. E., Birket, S., Field, T., Picking, W. D., Blocker, A. J., Galyov, E. E., Picking, W. L. & Lea, S. M. (2007). Self-chaperoning of the type III secretion system needle tip proteins IpaD and BipD. *J Biol Chem* **282**, 4035-44.
10. Quinaud, M., Chabert, J., Faudry, E., Neumann, E., Lemaire, D., Pastor, A., Elsen, S., Dessen, A. & Attree, I. (2005). The PscE-PscF-PscG complex controls type III secretion needle biogenesis in *Pseudomonas aeruginosa*. *J Biol Chem* **280**, 36293-300.
11. Sun, P., Tropea, J. E., Austin, B. P., Cherry, S. & Waugh, D. S. (2008). Structural characterization of the *Yersinia pestis* type III secretion system needle protein YscF in complex with its heterodimeric chaperone YscE/YscG. *J Mol Biol* **377**, 819-30.
12. Wang, Y., Ouellette, A. N., Egan, C. W., Rathinavelan, T., Im, W. & De Guzman, R. N. (2007). Differences in the electrostatic surfaces of the type III secretion needle proteins PrgI, BsaL, and MxiH. *J Mol Biol* **371**, 1304-14.
13. Zhang, L., Wang, Y., Picking, W. L., Picking, W. D. & De Guzman, R. N. (2006). Solution structure of monomeric BsaL, the type III secretion needle protein of *Burkholderia pseudomallei*. *J Mol Biol* **359**, 322-30.
14. Erskine, P. T., Knight, M. J., Ruaux, A., Mikolajek, H., Wong Fat Sang, N., Withers, J., Gill, R., Wood, S. P., Wood, M., Fox, G. C. & Cooper, J. B. (2006). High resolution structure of BipD: an invasion protein associated with the type III secretion system of *Burkholderia pseudomallei*. *J Mol Biol* **363**, 125-36.
15. Derewenda, U., Mateja, A., Devedjiev, Y., Routzahn, K. M., Evdokimov, A. G., Derewenda, Z. S. & Waugh, D. S. (2004). The structure of *Yersinia pestis* V-antigen, an essential virulence factor and mediator of immunity against plague. *Structure* **12**, 301-6.
16. Marlovits, T. C., Kubori, T., Sukhan, A., Thomas, D. R., Galan, J. E. & Unger, V. M. (2004). Structural insights into the assembly of the type III secretion needle complex. *Science* **306**, 1040-2.

17. Olive, A. J., Kenjale, R., Espina, M., Moore, D. S., Picking, W. L. & Picking, W. D. (2007). Bile salts stimulate recruitment of IpaB to the *Shigella flexneri* surface, where it colocalizes with IpaD at the tip of the type III secretion needle. *Infect Immun* **75**, 2626-9.
18. Stensrud, K. F., Adam, P. R., La Mar, C. D., Olive, A. J., Lushington, G. H., Sudharsan, R., Shelton, N. L., Givens, R. S., Picking, W. L. & Picking, W. D. (2008). Deoxycholate interacts with IpaD of *Shigella flexneri* in inducing the recruitment of IpaB to the type III secretion apparatus needle tip. *J Biol Chem* **283**, 18646-54.

Chapter 8: NMR Structure Determination of the Andes Hantavirus Nucleocapsid Protein Coiled-coil Domain

INTRODUCTION

Hantaviruses can cause two emerging infectious diseases known as the hantavirus cardiopulmonary syndrome (HCPS) and the hantavirus hemorrhagic fever with renal syndrome.[1] Annually, there are over 150,000 cases of hantaviral infections reported world wide.[2] Rodents are the primary reservoir of hantaviruses, and humans are normally infected by inhalation of aerosol contaminated with the excreta of infected rodents. In 1993, the first reported cases of HCPS in North America[3] was caused by a novel hantaviral species[4; 5], the Sin Nombre virus, and had an initial mortality rate of 78%. HCPS has since been reported throughout the United States with a current mortality rate of 35% when correctly diagnosed.[6] The major cause of HCPS in South America is the Andes virus, and person-to-person transmission of the Andes virus was reported in Argentina and Chile.[7] Hantaviruses are known to invade and replicate primarily in endothelial cells, including the endothelium of vascular tissues lining the heart, without any apparent cytotoxic effect.[8; 9; 10] Although the molecular events remain poorly understood, the decreased heart pumping capacity (leading to cardiogenic shock) in HCPS is not a direct result of cardiac cell death upon hantavirus infection.[11] Hence, the myocardial depression seen in HCPS is hypothesized to be mediated by an unknown factor, possibly related to the immune response to infection[12] or the release of a secondary signaling molecule involved in

inflammation like nitrous oxide, which was shown to be elevated during hantaviral infections.[13; 14; 15]

The genome of hantaviruses consists of three negative-stranded RNAs designated as small (S), medium (M), and large (L), which encode the nucleocapsid (N) protein, two integral membrane glycoproteins (G1and G2), and an RNA dependent RNA polymerase (L protein), respectively. G1 and G2 are glycoproteins that form heterodimers on the viral surface. In G1 and G2, the amino termini are glycosylated and form the ectodomains, followed by single pass transmembrane helices, then the carboxy termini or cytoplasmic tails project within the virions.[16] The G1 protein cytoplasmic tail was suggested to bind the viral RNP (ribonucleoprotein) and is critical for the viral assembly.[16] The RNA dependent RNA polymerase (L protein) is believed to bind three viral RNA segments and perform transcriptase and replicase function. The detailed knowledge on viral assembly and viral pathogenicity still remains unknown.

The N protein is highly immunogenic[17; 18] and elicits a strong immune response, which confers protection in mice.[19; 20; 21] It is highly conserved among hantavirus species and is the most abundant viral protein. It plays important roles in viral encapsidation, RNA packaging, and host-pathogen interaction.[22] The N protein binds to viral proteins[22], host proteins,[23; 24; 25; 26; 27; 28; 29] and viral RNA.[30; 31; 32; 33; 34] The self-association of the N protein into trimers was shown by gradient fractionation and chemical cross-linking.[35] Deletion mapping identified that regions at the N and C termini are important in N-N interaction,[35; 36; 37] and a model of N protein

trimerization was proposed by electron microscopy data to adopt the head-to-head and tail-to-tail association of the N-terminal and C-terminal domains, respectively.[36; 38] The N-terminal region in the Sin Nombre virus N protein (residues 3–73)[39] and the Tula virus (residues 1–77)[40] were predicted to form coiled coil domains. Recently, the structure of the N-terminal coiled coil domain (residues 1–75 and1–93) of the Sin Nombre virus was determined by crystallography.[41]

In addition to trimerization, the N protein was reported to also be involved in interaction with other proteins such as the viral polymerase and a viral glycoprotein[22] and host cellular proteins. The interaction between SUMO-1 (small ubiquitin-related modifier) and the SUMO-1 conjugating enzyme Ubc9[23; 25; 27] with N protein were reported and was required to localize the N protein in the preinuclear region.[27] Binding was mapped to the middle of the Hantaan virus N protein (residues 101-238) and a consensus motif of Ubc9 and SUMO-1 (with the sequence MAKE).[27] Similar results were observed with the middle region of Seoul virus N protein. Specifically residues 100-125 were critical in interacting with SUMO-1 and Ubc9.[25] Another protein that was reported to bind the N protein is the Daxx protein, a cellular protein primarily involved in protein-protein interactions in apoptosis and transcriptional co-repression. The interaction was mapped to the the C-terminal 57 residues of the Puumala virus N protein and the C-terminal 142 residues of Daxx protein.[26] The N protein was also reported to bind to soluble actin monomers and the insoluble microfilaments and this binding was important in viral assembly.[29] Lately, the

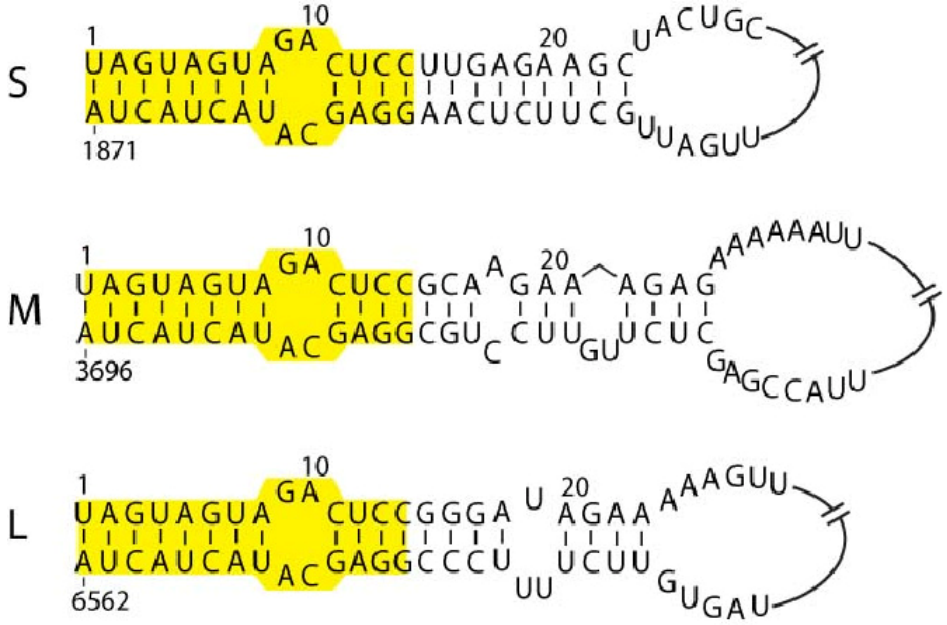

Figure 8-1. Predicted secondary structures of the three genomic segments (S, M and L) of Andes virus panhandles. The first 14-nucleotides are identical in the three segments (shaded).

association between N protein and a cellular protein MxA was shown to regulate the interferon-inducible immune response.[24]

Another function of the N protein is to bind the viral genome, which consists of three negative-strand RNA segments. N protein trimers bound with higher affinity and specificity to the viral RNA compared to monomers and dimers.[32] Synthetic peptides corresponding to the N protein central region (residues 175-217), which contains a significant number of conserved lysines and arginines, bound tightly to the viral RNA genome by gel shift assays.[34; 42; 43] The recombinant, His-tagged C-terminal 100 residues of the Puumala virus and Hantaan virus N protein bind the viral RNA genome nonspecifically, with higher affinity to double-stranded over signal-stranded RNA.[30] Futher identification showed N protein bound tightly at the 5' and 3' ends of the viral RNA genome, which have complementary sequences (Fig. 8-1) and were predicted to fold into "panhandle"-like secondary structures. The minimal 23-base pair viral panhandle of the Sin Nombre virus bound to the N protein with nanomolar affinity.[33]

In this chapter, we describe the solution structure of the N-terminal 1–74 residues of the Andes virus N protein, which forms a coiled coil domain. Further, we identified that the coiled coil contains distinct regions of positively and negatively charged surfaces involving conserved polar residues. We hypothesize that these regions are also important in N protein function. We used site-directed mutagenesis to alter the surface of the N protein and assayed for the subcellular localization of the N protein by immunocytochemistry. We used CD spectroscopy to confirm that

mutations did not alter the coiled coil structure of the $N^{1–74}$ (residues 1–74 of the N protein) domain. However, immunocytochemistry showed that despite the N protein being present throughout the cytoplasm, a monoclonal antibody only recognized the Arg22 and Lys26 mutants when nucleocapsids are bound to the Golgi, the site of viral assembly and maturation. We propose that the conserved surface residues Arg22 and Lys26 are important in the proper conformation or molecular recognition of the N protein. This chapter is adapted from the published paper by Wang et al.[44]

MATERIALS AND METHODS

Protein Expression and Purification of $N^{1–74}$

The $N^{1–74}$ domain of the Andes virus (strain 23) nucleocapsid protein was subcloned into pET-151 (Invitrogen), which appends a 33-residue His_6 tag and a TEV protease cleavage site at the N terminus. Isotopically ($^{15}N/^{15}N,^{13}C$) labeled protein was overexpressed in *Escherichia coli* BL21(DE3) (DNAY) grown in 1 liter of M9 minimal medium with 2g [^{13}C]glucose or/and 1g [^{15}N]ammonium chloride. The cells were grown at 37°C to A_{600} 0.8, induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and incubated overnight (16 h) in a 15°C shaker. The cells were harvested by centrifugation, resuspended in 30 ml of binding buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5 mM imidazole), and lysed by sonication. The recombinant protein was purified via His_6 tag as described. His_6 tag was removed using TEV protease from purified His_6- (TEV)- $N^{1–74}$. The TEV cleaved protein sample was further purified by the Ni^{2+} affinity column. All the purified protein

samples were concentrated using Amicon ultracel-3k if needed and then dialyzed into NMR buffer (10 mM sodium phosphate, pH 6.9, 10 mM NaCl) before use and the protein concentration was estimated using absorbance at 280 nm.

Mutagenesis of $N^{1–74}$

Site-specific mutations in the $N^{1–74}$ domain were introduced by PCR using the Stratagene QuikChange kit in two plasmids: (i) pET-151-$N^{1–74}$ (Amp^r), used to overexpress recombinant His-tagged $N^{1–74}$ in *E. coli*, and (ii) pcDNA3.1-AND-N (Amp^r), used to express full-length Andes virus N protein in a mammalian cell line for immunocytochemistry (see below). The mutations were confirmed by DNA sequencing.

NMR Spectroscopy

NMR data were acquired at 25°C using a Bruker Avance 800 MHz spectrometer equipped with a cryoprobe, processed with NMRPipe,[45] and analyzed with NMRView.[46] Backbone assignments were obtained from 2D ^1H-^{15}N HSQC[47] and 3D HNCA,[48] CBCA(CO)NH,[48] HNCACB,[49] and HNCO (Table 8-S1).[50] Secondary structures were identified from the C^α, C^β, C', and H^α chemical shifts.[51] Side chain assignments were obtained from 2D ^1H-^{13}C HMQC,[52] 3D HBHA(CO)NH,[53] and 3D ^{13}C-edited HMQC-NOESY (t_{mix} = 120 ms).[54] Nuclear Overhauser effect (NOE) cross-peaks were identified from 3D ^{15}N-edited NOESY-HSQC (t_{mix} = 120 ms)[55] and 3D ^{13}C-edited HMQC-NOESY (t_{mix} = 120 ms) (Table 8-S1).[54] Hydrogen-deuterium

exchange was performed by lyophilizing a 600-µl ^{15}N-labeled NMR sample and resuspending in 600 µl of 50% D_2O, 50% H_2O, followed by acquisition of six consecutive 20-min 2D ^{1}H-^{15}N HSQC spectra. Peak volumes were analyzed to identify residues with slower hydrogen-deuterium exchange rates.

Structure Calculation

NOE distance restraints were classified into upper bounds of 2.7, 3.5, 4.5, and 5.5 Å and lower bound of 1.8 Å based on peak volumes. Backbone dihedral angles in the α-helical regions were restrained to ϕ(-60 ± 20°) and φ(-40 ± 20°). Hydrogen bonding distance restraints were used for α-helical residues that showed slow hydrogen-deuterium exchange rates. Initial structures were generated by torsion angle dynamics in CYANA,[56] followed by molecular dynamics and simulated annealing in AMBER7,[57] first *in vacuo* and then with the generalized Born potential to account for the effect of solvent during structure calculation. Details of CYANA and AMBER structure calculation protocols have been described elsewhere.[58] Iterative cycles of AMBER calculations followed by refinement of NMR-derived restraints were performed until the structures converged with low restraint violations and good statistics in the Ramachandran plot. A family of 20 lowest energy structures was analyzed using PROCHECK,[59] and graphics were generated using PyMol.[60] The surface electrostatic potentials were calculated using APBS[61] and visualized in PyMol.[60]

CD Spectroscopy

N^{1-74} samples for CD spectroscopy contained 5–10 μM protein in buffer (25 μM Tris-HCl, pH 8, 3 μM EDTA, and 5 μM dithiothreitol). CD spectra were collected on a Jasco J-815 spectropolarimeter in triplicate. Wavelength scans were collected at 25°C at a scanning rate of 50 nm/min. Thermal denaturation scans at 222 nm were acquired with a temperature ramp rate of 1 °C/min to a final temperature of 80 °C, followed by cooling at 1°C/min to 25°C. The melting temperature (T_m) was determined from calculating the first derivative of thermal denaturation plots using the Jasco CD software.

Immunocytochemistry

Immunocytochemistry was performed as reported.[62] Briefly, Cos-7 cells (ATCC; # CRL-1651) were grown overnight in 24-well plates with coverslips at 37°C and 5% CO_2 in Dulbecco's modified Eagle's medium containing 10% fetal bovine serum. The cells at 80% confluence were transfected using Lipofectamine 2000 (Invitrogen) with 0.8 μg of pcDNA3.1-AND-N plasmid, which expresses full-length wild type or mutated N protein. At 48 h after transfection, the cells were washed with ice-cold phosphate-buffered saline and fixed at room temperature with methanol:acetone (3:1) for 10 min. The cells were incubated in 10 mM glycine for 30 min and permeabilized in phosphate-buffered saline with 0.1% Triton X-100 for 30 min. Permeabilized cells were incubated with antibodies for 60 min at room temperature and washed for 5 min three times with 0.3% Tween in phosphate-buffered saline after each incubation. Goat serum (10%) was used as a blocking agent.

Primary antibodies were of two sets: (i) rabbit polyclonal anti-hantavirus nucleocapsid (1:1000) (Immunology Consultants Laboratory; No. RSNV-55) and mouse monoclonal anti-hantavirus-nucleocapsid (1:1000) (Abcam; No. AB34757) or (ii) rabbit anti-Golgi matrix protein GM130 (1:200) (Calbiochem; No. CB1008) and mouse-anti-hantavirus nucleocapsid (1:1000). Secondary antibodies used were Alexa-Fluor-488 (1:1000) (Invitrogen; No. A11008) and Alexa-Fluor-594 (1:1000) (Invitrogen; No. A11005). Lastly, the cells were stained with 4',6-diamidino-2-phenylindole (Bio-Genex; No. CS2010-06), mounted on slides, and visualized at 60x on an Olympus FV1000 confocal microscope. The images were cropped and adjusted using Adobe Photoshop CS2.

RESULTS

NMR Structure Determination of N^{1-74}

The His_6-N^{1-74} expressed well in soluble form in *E. coli* and yielded an excellent 2D ^{1}H-^{15}N HSQC spectrum that showed distinct and well dispersed peaks (Fig. 8-2A). Nearly complete backbone assignments were obtained from 3D HNCA (Fig. 8-S1), CBCA(CO)NH, HNCACB (Fig. 8-S2), and ^{15}N-edited NOESY-HSQC (Fig. 8-S3). The histidine residues of the His tag were overlapped and could not be assigned unambiguously. The C_α, H_α, C_β, and C' secondary chemical shifts (Fig. 8-S4) showed that the first 33 residues, which were part of the His_6 tag, were in random coil orientation, and the native N^{1-74} sequence

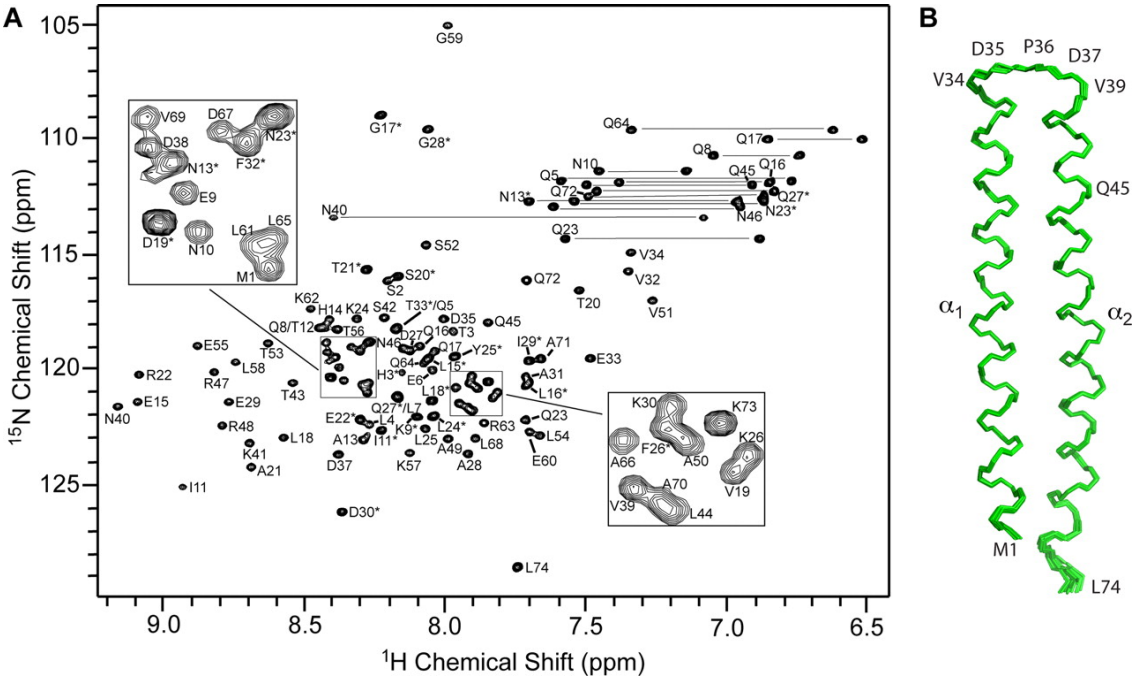

Figure 8-2. (A) Assigned ^{1}H-^{15}N HSQC spectrum of Andes virus N^{1-74} domain. The 33 N-terminal residues (shown with *) are part of the His-tag introduced by pET-151. The boxes show expansions of the crowded regions. **(B)** Superposition of 20 low energy NMR structures of the Andes virus N^{1-74} coiled coil domain. N^{1-74} forms two α-helices, M1-V34 and V39-L74. The anti-nucleocapsid monoclonal antibody used in immunocytochemistry below recognizes an epitope somewhere between residues 1-45.[44]

contained two α-helices.[51] The core domain structure was not affected by the flexible 33-residue His_6 tag (Fig. 8-S5). Side chain assignments were completed using 2D ^{1}H-^{13}C HMQC (Fig. 8-S6), 3D HBHA(CO)NH, and 3D ^{13}C-edited HMQC-NOESY. Manual analysis of 3D ^{15}N- and ^{13}C-edited NOESY spectra identified 1432 unambiguous interproton distance restraints. The NOE restraints together with 73 ϕ and 62 φ dihedral angle restraints and 38 hydrogen bond restraints (Table 8-1 and Fig. 8-S7) were used in structure calculation and refinement using CYANA and AMBER. The 20 lowest energy NMR structures of $N^{1–74}$ converged into a family of structures (Fig. 8-2B) with low restraint violations and good Ramachandran plot statistics (Table 8-1).

The $N^{1–74}$ Coiled Coil Domain

$N^{1–74}$ forms two well defined α-helices (α_1, Met1–Val34; α_2, Val39–Leu74) that are connected by an ordered acidic loop (Asp35-Pro36-Asp37-Asp38) (Fig. 8-2B). The two helices are intertwined into a coiled coil, and the helix α_1-α_2 interface is lined with hydrophobic amino acids positioned in every seventh residue on helix α_1 (Leu4, Ile11, Leu18, Leu25, and Val32) and helix α_2 (Leu44, Val51, Leu58, and Leu65) (Fig. 8-3A). This heptad repeat of hydrophobic residues is a hallmark of coiled coils and is highly conserved among hantaviruses.[40] Together with Pro36, the heptad repeats of leucines, isoleucines, and valines form the hydrophobic core that stabilize the structure of the coiled coil (Fig. 8-4A). On the same face of the hydrophobic heptad, there is another seven-residue repeat, in this case, composed of

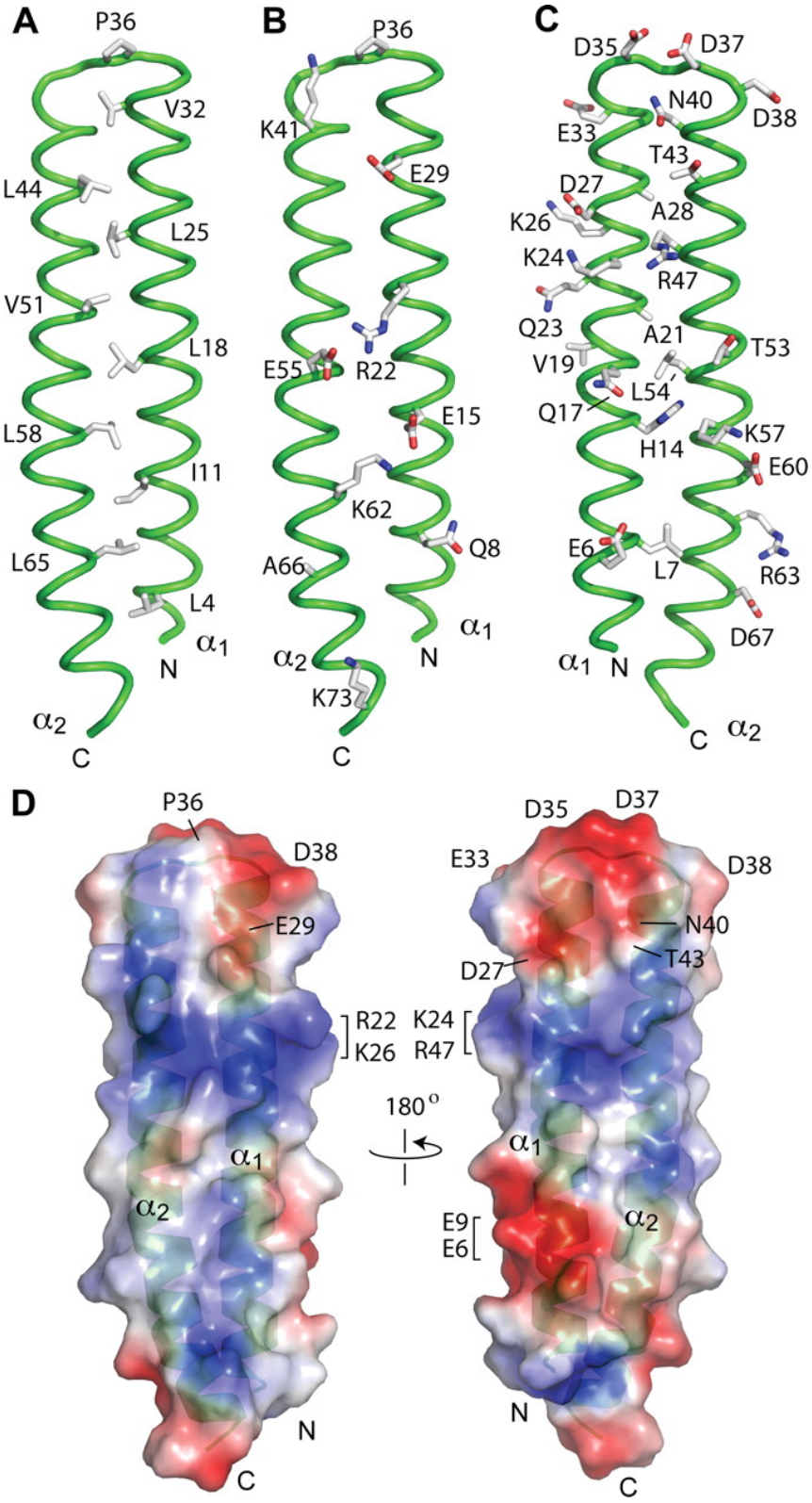
A
P36
V32
L44
L25
V51
L18
L58
I11
L65
L4
α1
N
α2
C
B
P36
K41
E29
E55
R22
E15
K62
Q8
A66
α1
α2
N
K73
C
C
D35
D37
N40
D38
E33
T43
D27
A28
K26
K24
R47
Q23
A21
T53
V19
L54
Q17
K57
H14
E60
E6
L7
R63
D67
α1
N
C
α2
D
P36
D38
E29
R22
K26
α1
α2
N
C
180°
D35
D37
E33
D38
N40
T43
D27
K24
R47
α1
E9
E6
α2
N
C

Figure 8-3. (**A**) Heptad repeats of conserved hydrophobic residues form the interface of the helix α_1 and α_2 that stabilize the coiled coil domain. (**B**) There is also a heptad repeat of polar residues, some of which (Arg22–Glu55 and Glu15–Lys62) form salt bridges that contribute in stabilizing the coiled coil. (**C**) There are many highly conserved residues that point away from the coiled coil and thus are not involved in stabilizing the coiled coil. (**D**) Electrostatic surface potential map of N^{1-74}. The orientation of the *left panel* is identical to that in *A* and is rotated 180° from the *right panel*, which is identical in the orientation of **C**. Conserved surface residues forming the acidic (*red*) and basic (*blue*) surfaces are indicated. Point mutations of Arg22 and Lys26 had a dramatic effect on the antibody recognition of the N protein *in vivo*.[44]

Table 8-1. Structural statistics for 20 NMR structures of Andes virus N^{1-74} coiled coil domain.[44]

Total unambiguous distance restraints	1432
Intraresidue (i,i)	92
Sequential (i, i+1)	361
Medium Range (2<=|i-j|<=4)	700
Long Range (|i-j| > 4)	279
Hydrogen bond restraints (19 hydrogen bonds)	38
Total dihedral angle restraints	135
Phi	73
Psi	62
RMS difference from mean structure	
Backbone atoms (N,C^{α},C') (Å)	0.21
All heavy atoms (C,N,O) (Å)	0.53
Violation analysis	
Max distance violation (Å)	0.27
Max dihedral angle violation (°)	3.2
Energies	
mean AMBER energy (kcal mol^{-1})	-3861
mean restraint energy (kcal mol^{-1})	117
Ramachandran plot	
Most favorable region	91.4%
Additionally allowed regions	8.2%
Generously allowed regions	0.4%
Disallowed regions	0.1%

polar residues on helix α_1 (Gln8, Glu15, Arg22, and Glu29) (Fig. 8-3B), which are invariant among the hantaviruses (Fig. 8-S8). Helix α_2 also contains a polar heptad, with more residue variability at positions 41 (Lys), 48 (Arg/Gln/Glu), 55 (Glu/Gln), and 62 (Lys/Arg). These polar residues form two conserved salt bridges between helix α_1-α_2 (Glu15–Lys62 and Arg22–Glu55) (Fig. 8-3B). Gln8 and Lys41 are surface-exposed and do not form obvious salt bridges; however, they are invariant among the hantaviruses, suggesting an unknown function.

In addition to the conserved heptad repeats mentioned above, there are other highly conserved residues whose side chains are pointed toward the helix α_1-α_2 interface. These residues are nonpolar (Leu7 and Leu54), aromatic (His14), polar (Gln17, Asn40, and Thr43), or charged (Glu6, Glu15, Glu29, Lys41, Arg47, and Lys57), and their side chains are pointed toward the helix α_1-α_2 interface (Fig. 8-3C). The polar and charged residues in this group do not participate in any salt bridge or hydrogen bonding contacts; however, their polar moieties are pointed toward the surface of the coiled coil, whereas the aliphatic portion of their side chains are involved in hydrophobic interaction that contribute to the stabilization of the hydrophobic core. The methyl groups of two invariant alanines, Ala21 and Ala28, in helix α_1 (Fig. 8-3C) are oriented toward the helix α_1-α_2 interface but do not contact any other residues on helix α_2, indicating that small side chains are required in those positions.

Conserved Surface Residues

A striking feature of the N^{1-74} coiled coil is the presence of large numbers of highly conserved residues whose side chains are pointed away from the coiled coil. These residues are nonpolar (Ala66 and Val19), polar (Gln8 and Gln23), basic (Lys24, Lys26, Arg63, and Lys73), and acidic (Asp27, Glu33, Asp35, Asp37, Asp38, Glu60, and Asp67) (Fig. 8-3B and 3C). These polar residues are identical (Gln8, Gln23, Lys24, Glu33, Asp35, Asp37, Arg63, and Lys73) or semi-identical (basic residues in position 26 and acidic residues in positions 27, 38, and 60) among hantaviruses (Fig. 8-S8). Further, many residues in this group are clustered together on the surface. The first cluster (Gln23, Lys24, and Lys26 together with Arg47 and Arg22 discussed in the preceding paragraph) forms a basic surface (Fig. 8-3D), and the second cluster (Asp27, Glu33, Asp35, Asp37, and Asp38) forms an acidic surface (Fig. 8-3D). We mutagenized many residues in this group (see below) to test the hypothesis that these residues are important in molecular recognition rather than in stabilizing the coiled coil structure.

Electrostatic Surface of N^{1-74}

The N^{1-74} domain is acidic (theoretical pI of 5.8), and the surface electrostatic potential map of N^{1-74} shows distinct regions of negatively charged (*red*) and positively charged (*blue*) surfaces (Fig. 8-3D). The tip of the coiled coil, where the loop connecting the two helices is located, is negatively charged (Fig. 8-3D) because of clustering of conserved acidic residues (Asp27, Glu29, Glu33, Asp35, Asp37, and Asp38) and polar residues (Asn40 and Thr43). Although the N^{1-74} domain is acidic,

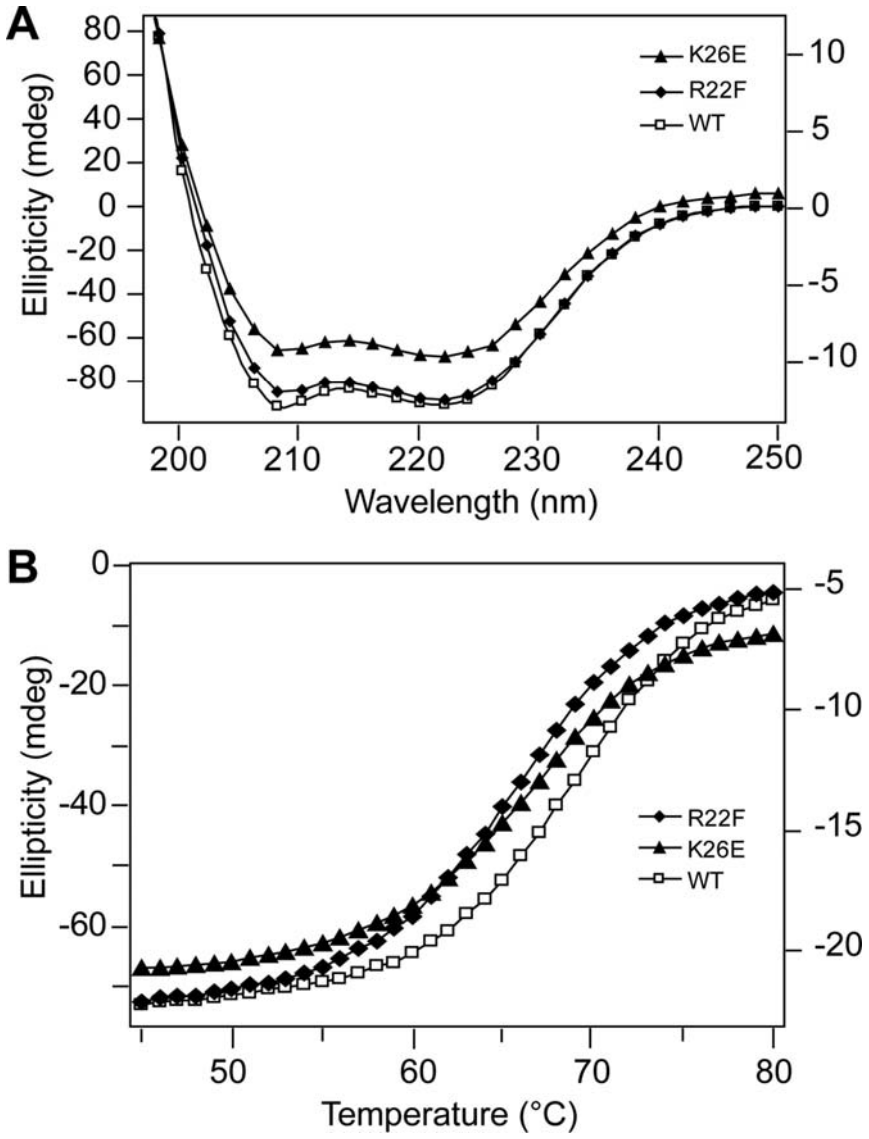

Figure 8-4. (**A**) CD spectra of N^{1-74} wild type (*WT*) and point mutants (K26E and R22F) showing the characteristic α-helical dips at 208 and 222 nm. All other N^{1-74} point mutants (listed in Table 8-2) showed similar α-helical CD spectra. (**B**) CD thermal denaturation curves, monitored at 222 nm, of wild type N^{1-74} and two point mutants, K26E and R22F. The rest of the point mutants showed similar thermal denaturation curves. The ellipticity scales on the *y* axes are shown on *left* (wild type and R22F) and *right* (K26E).[44] (Data courtesy of Chet W. Egan)

Table 8-2. Melting temperatures (Tm) and ellipticity (θ) ration at 222 and 208 nm of N^{1-74} .[44] (Data courtesy of Chet W. Egan)

N^{1-74}	T_m (°C)	$\theta_{222}/\theta_{208}$	Change in surface property
D38R	64.3 ± 0.01	1.06	acidic to basic
E33K	64.4 ± 0.02	1.05	acidic to basic
P36G	66.0 ± 0.1	0.99	increased loop flexibility
R47E	66.0 ± 0.03	1.07	basic to acidic
D38L	66.2 ± 0.01	1.03	acidic to nonpolar
R22F	66.4 ± 0.1	1.07	basic to bulky nonpolar
K26E	67.0 ± 0.02	1.04	basic to acidic
E33L	67.4 ± 0.01	1.07	acidic to nonpolar
Q23L	68.5 ± 0.03	1.01	polar to nonpolar
R22M	69.3 ± 1.2	1.06	basic to nonpolar
WT	69.4 ± 0.1	0.99	no change
K24A	70.6 ± 0.01	1.04	basic to small nonpolar
R47A	74.0 ± 0.02	1.02	basic to small nonpolar

there are conserved basic residues (Arg22, Lys24, Lys26 and Arg47) that form a positively charged surface just below the negatively charged tip (Fig. 8-3D). Point mutations in this positively charged surface have a dramatic effect on the antibody recognition of the N protein *in vivo* (see below).

In addition, there is a smaller negatively charged surface formed by Glu9 and Glu6 (Fig. 8-3D). Residue 9 could be acidic (Glu or Asp) or basic (Arg or Lys). Residue 9 is acidic among American hantaviruses (which cause the cardiopulmonary syndrome) and Old World hantaviruses that are nonpathogenic or cause a milder form of hemorrhagic fever with renal syndrome. Residue 9 is basic among Old World hantaviruses that causes the severe form of hemorrhagic fever with renal syndrome.

CD Spectroscopy of $N^{1–74}$

Point mutations were introduced in the basic (Arg22, Lys24, Lys26, and Arg47) and acidic (Glu33 and Asp38) surfaces. In addition, we mutated Gln23, which is near the basic region, and Pro36, which is near the acidic region. These residues are surface-exposed (Fig. 8-3D) and are nearly invariant among hantaviruses (Fig. 8-S8). CD spectroscopy was used to assess the folding and stability of $N^{1–74}$ mutants. Wild type and point mutants showed nearly identical CD spectra (Fig. 8-4A), indicating that the α-helical structure of $N^{1–74}$ was preserved. In addition, the ratio of ellipticity at 222 and 208 nm can be used to characterize α-helices. A $\theta_{222}/\theta_{208}$ ratio of ~1.0 indicates α-helices with extensive interhelical contacts as in coiled coils and helical bundles, whereas a $\theta_{222}/\theta_{208}$ ratio of ~0.8 indicates extended α-helices with little

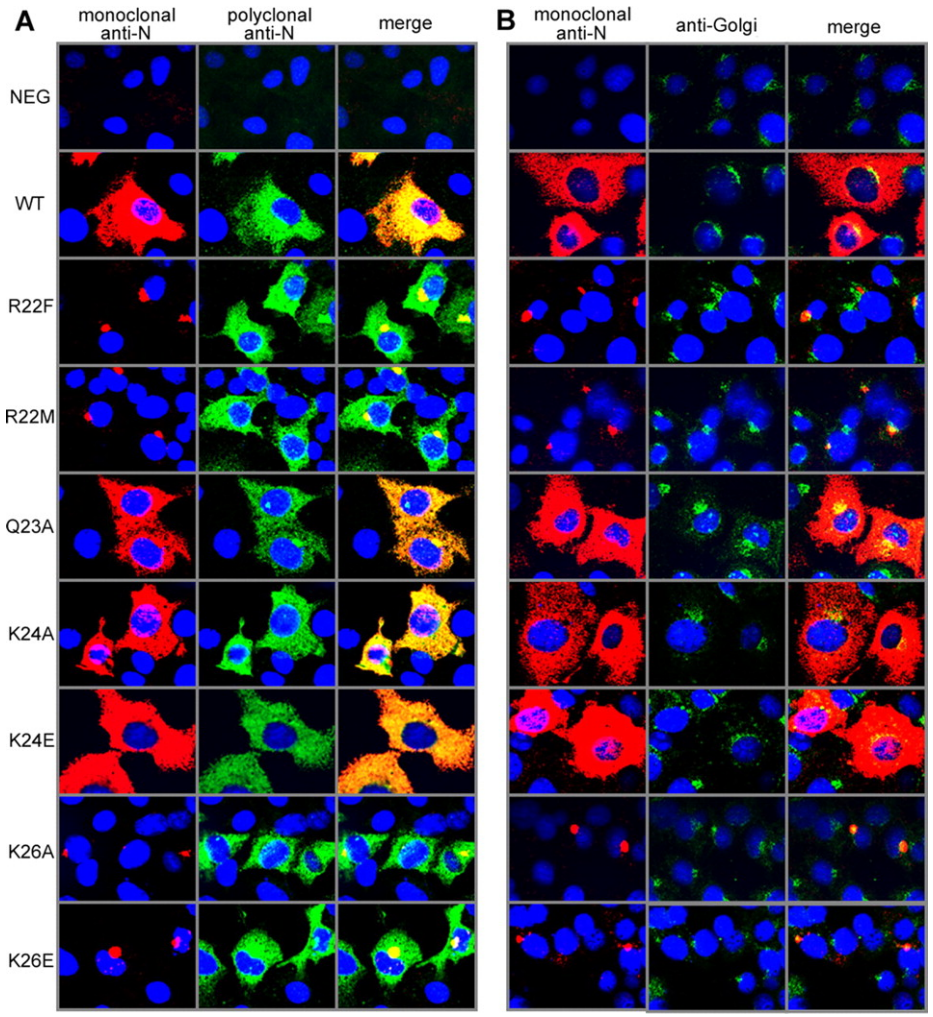

Figure 8-5. Immunocytochemistry of full-length N protein with point mutations in the $N^{1–74}$ coiled coil domain. Cos-7 cells were transfected with a plasmid expressing Andes virus N protein. Two days after transfection, the cells were fixed for immunofluorescence microscopy and double labeled with monoclonal (*red*) and polyclonal (*green*) anti-nucleocapsid antibodies (**A**) and monoclonal anti-nucleocapsid antibody (*red*) and anti-Golgi antibody (*green*) (**B**). The cell nuclei were stained *blue* using 4',6-diamidino-2-phenylindole. Point mutations in Arg22 and Lys26 showed a dramatic difference in the monoclonal antibody recognition of Golgi-associated N protein, suggesting that the conformation or molecular interaction (or both) of the N protein is different when it is in the cytoplasm or when it is associated with the Golgi. *WT*, wild type.[44] (Data courtesy of Daniel M. Boudreaux from Dr. Stephen C. St. Jeor, University of Nevada, Reno)

interhelical contacts.[63; 64; 65] All N^{1-74} constructs have a $\theta_{222}/\theta_{208}$ ratio higher than 0.9 (Table 8-2), suggesting that all mutants have the intact coiled coil structure. Further insight was provided by acquiring the CD melting temperatures (Fig. 8-4B and Table 8-2). Compared with wild type N^{1-74}, the majority of mutants showed lower T_m, with D38R having the lowest value, whereas two mutants (K24A and R47A) showed higher T_m (Table 8-2). Nevertheless, all mutations were within ±5°C of wild type T_m (Table 8-2), indicating that the mutations did not drastically alter the thermal stability of N^{1-74}. Thus, the point mutations maintained the structural integrity of the N^{1-74} coiled-coil.

Immunocytochemistry of N Protein

Hantaviruses are believed to mature intracellularly; specifically, in the Golgi complex.[66] During infection, the N protein was shown to localize cytoplasmically in the endoplasmic reticulum-Golgi intermediate compartment, presumably as they traffic from the endoplasmic reticulum to the Golgi.[28] In addition, immunofluorescence of Cos-7 cells transfected with the N protein alone showed a granular pattern of staining in the perinuclear region,[38; 40] suggesting colocalization with the Golgi. To test our hypothesis that the conserved surface residues of N^{1-74} are important in molecular interaction, we introduced point mutations designed to keep the N^{1-74} coiled-coil domain intact while altering only specific surface residues and transfected full-length N protein in mammalian cells to observe the subcellular localization of the N protein. Immunocytochemistry was performed by Daniel M.

Bourdreaux in the laboratory of Dr. Stephen C. St. Joer at the University of Nevada, Reno. Two types of anti-nucleocapsid antibodies, rabbit polyclonal and mouse monoclonal antibodies were used. The polyclonal antibody detected that wild type N and mutants (Arg22, Gln23, Lys24, and Lys26) were located throughout the cytoplasm (Fig. 8-5A). The monoclonal antibody also detected wild type N and the Gln23 and Lys24 mutants throughout the cytoplasm in a similar pattern of staining as the polyclonal antibody (Fig. 8-5A). However, the monoclonal antibody showed a dramatic difference between the recognition of wild type N and the Arg22 and Lys26 mutants (Fig. 8-5A). Using the monoclonal antibody, Arg22 and Lys26 mutants were observed in a compact location lateral to the nucleus (Fig. 8-5A). To further define the subcellular localization of these N mutants, a Golgi-specific antibody (targeting the Golgi matrix protein GM130) was used (Fig. 8-5B). The Arg22 and Lys26 mutants were only detected by the monoclonal antibody when the N protein colocalized with the Golgi (Fig. 8-5B); however, these mutants were also present throughout the cytoplasm as shown by the polyclonal antibody (Fig. 8-5A). Thus, for the Arg22 and Lys26 mutants, the monoclonal antibody was able to distinguish between two populations of the N protein based on its subcellular localization in the cytoplasm or in the Golgi, the site of viral assembly and maturation.[66] Other mutants (Glu33, Asp35, Pro36, Asp37, Asp38, and Arg47) did not show this localization-dependent antibody recognition (Fig. 8-S9).

DISSCUSSION

The NMR structure of the Andes virus $N^{1–74}$ domain is similar to the recent crystal structure of the Sin Nombre virus nucleocapsid protein N-terminal coiled-coil ($N^{1–75}$).[41] The C_α root mean square deviation (RMSD) between the two structures is 1.3 Å. The crystal structure determination of the N protein addressed the issue of the trimerization of the N protein[41] because earlier models suggested the trimerization of the nucleocapsid N-terminal domain.[35; 36; 38; 39] A proposed model of N protein trimerization involves, first, the association of three N-terminal domains, followed by the association of three C-terminal domain.[40] However, crystallography revealed that the Sin Nombre nucleocapsid N-terminal domain was monomeric and formed a coiled-coil structure, and conserved hydrophobic residues participate in helix-helix interaction that stabilize the coiled-coil.[41] Our NMR structure of the Andes virus $N^{1–74}$ supports the crystallographic results; even at 1.4 mM, $N^{1–74}$ remained monomeric in solution. Our results, however, do not preclude the trimerization of full-length N protein *in vivo* by another mechanism.

A feature of the $N^{1–74}$ domain that had not been addressed in the literature is the role of many conserved polar residues whose side chains are pointed away from the coiled-coil. Furthermore, the majority of these surface-exposed residues are not involved in polar interactions (Fig. 8-3D). Point mutations of these polar residues maintained the structural integrity and high thermal stability of the coiled-coil (Fig. 8-4 and Table 8-2). For example, Arg22, which forms a salt bridge with a conserved residue Glu^{55}, can be mutated (R22F or R22M) without disrupting the coiledc-coil

structure of N^{1-74} (Table 8-2). R22F, which replaced arginine with a bulkier aromatic side chain, decreased the overall melting temperature by ~3°C (Table 8-2). This change is likely attributed to increased steric clash between phenylalanine and Glu^{55}. However, the observation that R22M melts at a temperature comparable with that of wild type suggests that the salt bridge between Arg22 and Glu55 does not play a significant role in helix-helix interaction and that hydrophobic interaction is the major force stabilizing the coiled-coil. A mutation in a nonpolar residue, Pro36, which is at the turn connecting the two α-helices of the coiled-coil, had a T_m approximately four degrees lower than wild type, which is consistent with a mutation that increases the number of conformations available at the Pro36 turn and destabilizes the overall protein structure by uncoupling the helix-helix interaction. Nevertheless, all of the point mutations of the conserved surface residues maintained the coiled-coil structure of N^{1-74} (Fig. 8-4 and Table 8-2).

Thus, there is no compelling structural reason for the high sequence conservation of surface residues. Furthermore, these polar residues are clustered together on the surface of the N^{1-74} domain and form distinct positively and negatively charged regions (Fig. 8-3D). We hypothesize that the reason for the clustering of conserved polar residues on the surface of N^{1-74} is that they are sites of molecular recognition involved in the proper function of the N protein. Our mutagenesis and immunocytochemistry data suggest that point mutations in this group had a dramatic effect on the antibody recognition of the N protein with respect to its subcellular localization (Fig. 8-5).

During infection, nucleocapsids are trafficked to the cytoplasm[28] to assemble into mature virions.[66] Mammalian cells transfected with the N protein alone show a granular pattern of immunofluorescence.[38; 40] This localization pattern is thought to be necessary for the nucleocapsid to perform its many functions in the establishment of an effective infection.[28] We questioned whether the conserved polar surface residues in the coiled-coil domain are important in the proper functioning of the N protein and reasoned that defects in the conformation or molecular recognition of the N protein will be manifested in the antibody recognition of the N protein in the context of its subcellular localization. CD spectroscopy confirmed that the mutant forms of N^{1-74} maintained the structural integrity of the coiled-coil structure (Fig. 8-4 and Table 8-2); thus, the mutations altered only the surface property of the N protein.

Immunocytochemistry (Fig. 8-5) indicates that mutations in a conserved basic surface formed by Arg22 and Lys26 show monoclonal antibody recognition depending on the subcellular localization of the N protein. Polyclonal antibodies show that Arg22 and Lys26 mutants are present in the cytoplasm and Golgi; however, only Golgi-associated mutant nucleocapsids are detected by the monoclonal antibody (Fig. 8-5). Mutation of Arg22 or Lys26 changes the presentation of the N-terminal coiled-coil to the monoclonal antibody. This change is dependent on the subcellular localization of the N protein.

There are two possible scenarios that could account for this differential monoclonal antibody recognition of the Arg22 and Lys26 mutants. First, there may be a difference in the conformation of the N-terminal coiled-coil depending on whether

the N protein is localized in the cytoplasm or in the Golgi, and this conformational change upon binding to the Golgi exposes the epitope, which is somewhere between residues 1–45 (comprising helix α_1, the interhelical loop, and part of helix α_2 of the N^{1-74} coiled-coil; Fig. 8-2*B*), thereby allowing the monoclonal antibody to recognize the N protein associated with the Golgi. Second, the epitope may be masked differently by molecular interactions when the N protein is localized in the cytoplasm or in the Golgi. In addition to self-association, several host proteins such as SUMO-1,[23; 25; 27] Ubc9,[25; 27] Daxx[26] actin,[29] microtubules,[28] and MxA[24] were reported to bind the N protein. Binding of the N protein with SUMO-1 and Ubc9 was required for localization of the N protein in the perinuclear region.[23; 27] Furthermore, because the N protein is not known to be a membrane protein, its localization in the Golgi must involve interaction with a Golgi-associated protein. Any of these molecular interactions could potentially alter the epitope presentation of the N^{1-74} coiled-coil and needs to be experimentally verified.

In summary, our structural results revealed that the highly conserved polar residues in the N-terminal coiled-coil domain of the hantavirus nucleocapsid protein form distinct acidic and basic surfaces, and point mutations of the conserved basic surface formed by Arg22 and Lys26 allowed a monoclonal antibody to distinguish between two populations of the N protein based on its subcellular localization. Thus, in the Arg22 or Lys26 mutants, the conformation or molecular interaction of the N protein is different when it is in the cytoplasm or in the Golgi, the site of viral assembly and maturation. It is worth determining whether the Arg22 or Lys26

mutants affect the N protein trimerization or even the viral assembly. Similar to N-terminal, C-terminal of N protein was also mapped to be important for trimerization.[38] Further information on the atomic structure of C-terminal region and how the C-terminal is involved in the protein-protein interaction will be interesting to pursuit.

ACKNOWLEDGEMENT

This work is done by the collaboration with Dr. Stephen C. St. Jeor's lab at University of Nevada, Reno. The plasmids pET-151-$N^{1–74}$ and pcDNA3.1-AND-N were constructed in Dr. Stephen C. St. Jeor's laboratory. And Daniel M. Boudreaux performed the immunocytochemistry experiments of N protein mutans in Dr. Stephen C. St. Joer's lab. Chet W. Egan and Nan Wang in De Guzman's lab helped to purified some of the N^{1-74} mutant proteins and Chet W. Egan acquired the CD spectroscopy data of those mutants.

REFERENCES

1. Schmaljohn, C. & Hjelle, B. (1997). Hantaviruses: A Global Disease Problem. *Emerg. Infect. Dis.* **3**, 95-104.
2. Khaiboullina, S. F., Morzunov, S. P. & St Jeor, S. C. (2005). Hantaviruses: molecular biology, evolution and pathogenesis. *Curr Mol Med* **5**, 773-90.
3. Prevention, C. f. D. C. a. (1993). Outbreak of acute illness-Southwestern United States. *Morbid. Mortal. Weekly Rep.* **42**, 421-443.
4. Hjelle, B., Jenison, S., Torrez-Martinez, N., Yamada, T., Nolte, K., Zumwalt, R., MacInnes, K. & Myers, G. (1994). A novel hantavirus associated with an outbreak of fatal respiratory disease in the southwestern United States: evolutionary relationships to known hantaviruses. *J Virol* **68**, 592-6.
5. Nichol, S. T., Spiropoulou, C. F., Morzunov, S., Rollin, P. E., Ksiazek, T. G., Feldmann, H., Sanchez, A., Childs, J., Zaki, S. & Peters, C. J. (1993). Genetic

identification of a hantavirus associated with an outbreak of acute respiratory illness. *Science* **262**, 914-7.
6. Mertz, G. J., Hjelle, B., Crowley, M., Iwamoto, G., Tomicic, V. & Vial, P. A. (2006). Diagnosis and treatment of new world hantavirus infections. *Curr Opin Infect Dis* **19**, 437-42.
7. Padula, P. J., Edelstein, A., Miguel, S. D., Lopez, N. M., Rossi, C. M. & Rabinovich, R. D. (1998). Hantavirus pulmonary syndrome outbreak in Argentina: molecular evidence for person-to-person transmission of Andes virus. *Virology* **241**, 323-30.
8. Nolte, K. B., Feddersen, R. M., Foucar, K., Zaki, S. R., Koster, F. T., Madar, D., Merlin, T. L., McFeeley, P. J., Umland, E. T. & Zumwalt, R. E. (1995). Hantavirus pulmonary syndrome in the United States: a pathological description of a disease caused by a new agent. *Hum Pathol* **26**, 110-20.
9. Pensiero, M. N., Sharefkin, J. B., Dieffenbach, C. W. & Hay, J. (1992). Hantaan virus infection of human endothelial cells. *J Virol* **66**, 5929-36.
10. Zaki, S. R., Greer, P. W., Coffield, L. M., Goldsmith, C. S., Nolte, K. B., Foucar, K., Feddersen, R. M., Zumwalt, R. E., Miller, G. L., Khan, A. S. & et al. (1995). Hantavirus pulmonary syndrome. Pathogenesis of an emerging infectious disease. *Am J Pathol* **146**, 552-79.
11. Peters, C. J. & Zaki, S. R. (2002). Role of the endothelium in viral hemorrhagic fevers. *Crit Care Med* **30**, S268-73.
12. Kraus, A. A., Raftery, M. J., Giese, T., Ulrich, R., Zawatzky, R., Hippenstiel, S., Suttorp, N., Kruger, D. H. & Schonrich, G. (2004). Differential antiviral response of endothelial cells after infection with pathogenic and nonpathogenic hantaviruses. *J Virol* **78**, 6143-50.
13. Davis, I. C., Zajac, A. J., Nolte, K. B., Botten, J., Hjelle, B. & Matalon, S. (2002). Elevated generation of reactive oxygen/nitrogen species in hantavirus cardiopulmonary syndrome. *J Virol* **76**, 8347-59.
14. Klingstrom, J., Akerstrom, S., Hardestam, J., Stoltz, M., Simon, M., Falk, K. I., Mirazimi, A., Rottenberg, M. & Lundkvist, A. (2006). Nitric oxide and peroxynitrite have different antiviral effects against hantavirus replication and free mature virions. *Eur J Immunol* **36**, 2649-57.
15. Klingstrom, J., Hardestam, J. & Lundkvist, A. (2006). Dobrava, but not Saaremaa, hantavirus is lethal and induces nitric oxide production in suckling mice. *Microbes Infect* **8**, 728-37.
16. Estrada, D. F., Boudreaux, D. M., Zhong, D., St Jeor, S. C. & De Guzman, R. N. (2009). The hantavirus glycoprotein G1 tail contains a dual CCHC-type classical zinc fingers. *J Biol Chem*.
17. Gott, P., Zoller, L., Darai, G. & Bautz, E. K. (1997). A Major Antigenic Domain of Hantaviruses is Located on the Aminoproximal Site of the Viral Nucleocapsid Protein. *Virus Genes* **14**, 31-40.
18. Lundkvist, A., Meisel, H., Koletzki, D., Lankinen, H., Cifire, F., Geldmacher, A., Sibold, C., Gott, P., Vaheri, A., Kruger, D. H. & Ulrich, R. (2002).

Mapping of B-cell epitopes in the nucleocapsid protein of Puumala hantavirus. *Viral Immunol* **15**, 177-92.

19. Geldmacher, A., Skrastina, D., Borisova, G., Petrovskis, I., Kruger, D. H., Pumpens, P. & Ulrich, R. (2005). A hantavirus nucleocapsid protein segment exposed on hepatitis B virus core particles is highly immunogenic in mice when applied without adjuvants or in the presence of pre-existing anti-core antibodies. *Vaccine* **23**, 3973-83.
20. Geldmacher, A., Skrastina, D., Petrovskis, I., Borisova, G., Berriman, J. A., Roseman, A. M., Crowther, R. A., Fischer, J., Musema, S., Gelderblom, H. R., Lundkvist, A., Renhofa, R., Ose, V., Kruger, D. H., Pumpens, P. & Ulrich, R. (2004). An amino-terminal segment of hantavirus nucleocapsid protein presented on hepatitis B virus core particles induces a strong and highly cross-reactive antibody response in mice. *Virology* **323**, 108-19.
21. Maes, P., Keyaerts, E., Bonnet, V., Clement, J., Avsic-Zupanc, T., Robert, A. & Van Ranst, M. (2006). Truncated recombinant Dobrava hantavirus nucleocapsid proteins induce strong, long-lasting immune responses in mice. *Intervirology* **49**, 253-60.
22. Kaukinen, P., Vaheri, A. & Plyusnin, A. (2005). Hantavirus nucleocapsid protein: a multifunctional molecule with both housekeeping and ambassadorial duties. *Arch Virol* **150**, 1693-713.
23. Kaukinen, P., Vaheri, A. & Plyusnin, A. (2003). Non-covalent interaction between nucleocapsid protein of Tula hantavirus and small ubiquitin-related modifier-1, SUMO-1. *Virus Res* **92**, 37-45.
24. Khaiboullina, S. F., Rizvanov, A. A., Deyde, V. M. & St Jeor, S. C. (2005). Andes virus stimulates interferon-inducible MxA protein expression in endothelial cells. *J Med Virol* **75**, 267-75.
25. Lee, B. H., Yoshimatsu, K., Maeda, A., Ochiai, K., Morimatsu, M., Araki, K., Ogino, M., Morikawa, S. & Arikawa, J. (2003). Association of the nucleocapsid protein of the Seoul and Hantaan hantaviruses with small ubiquitin-like modifier-1-related molecules. *Virus Res* **98**, 83-91.
26. Li, X. D., Makela, T. P., Guo, D., Soliymani, R., Koistinen, V., Vapalahti, O., Vaheri, A. & Lankinen, H. (2002). Hantavirus nucleocapsid protein interacts with the Fas-mediated apoptosis enhancer Daxx. *J Gen Virol* **83**, 759-66.
27. Maeda, A., Lee, B. H., Yoshimatsu, K., Saijo, M., Kurane, I., Arikawa, J. & Morikawa, S. (2003). The intracellular association of the nucleocapsid protein (NP) of hantaan virus (HTNV) with small ubiquitin-like modifier-1 (SUMO-1) conjugating enzyme 9 (Ubc9). *Virology* **305**, 288-97.
28. Ramanathan, H. N., Chung, D. H., Plane, S. J., Sztul, E., Chu, Y. K., Guttieri, M. C., McDowell, M., Ali, G. & Jonsson, C. B. (2007). Dynein-dependent transport of the hantaan virus nucleocapsid protein to the endoplasmic reticulum-Golgi intermediate compartment. *J Virol* **81**, 8634-47.
29. Ravkov, E. V., Nichol, S. T., Peters, C. J. & Compans, R. W. (1998). Role of actin microfilaments in Black Creek Canal virus morphogenesis. *J Virol* **72**, 2865-70.

30. Gott, P., Stohwasser, R., Schnitzler, P., Darai, G. & Bautz, E. K. (1993). RNA binding of recombinant nucleocapsid proteins of hantaviruses. *Virology* **194**, 332-7.
31. Mir, M. A., Brown, B., Hjelle, B., Duran, W. A. & Panganiban, A. T. (2006). Hantavirus N protein exhibits genus-specific recognition of the viral RNA panhandle. *J Virol* **80**, 11283-92.
32. Mir, M. A. & Panganiban, A. T. (2004). Trimeric hantavirus nucleocapsid protein binds specifically to the viral RNA panhandle. *J Virol* **78**, 8281-8.
33. Mir, M. A. & Panganiban, A. T. (2005). The hantavirus nucleocapsid protein recognizes specific features of the viral RNA panhandle and is altered in conformation upon RNA binding. *J Virol* **79**, 1824-35.
34. Severson, W., Partin, L., Schmaljohn, C. S. & Jonsson, C. B. (1999). Characterization of the Hantaan nucleocapsid protein-ribonucleic acid interaction. *J Biol Chem* **274**, 33732-9.
35. Alfadhli, A., Love, Z., Arvidson, B., Seeds, J., Willey, J. & Barklis, E. (2001). Hantavirus nucleocapsid protein oligomerization. *J Virol* **75**, 2019-23.
36. Kaukinen, P., Vaheri, A. & Plyusnin, A. (2003). Mapping of the regions involved in homotypic interactions of Tula hantavirus N protein. *J Virol* **77**, 10910-6.
37. Yoshimatsu, K., Lee, B. H., Araki, K., Morimatsu, M., Ogino, M., Ebihara, H. & Arikawa, J. (2003). The multimerization of hantavirus nucleocapsid protein depends on type-specific epitopes. *J Virol* **77**, 943-52.
38. Kaukinen, P., Kumar, V., Tulimaki, K., Engelhardt, P., Vaheri, A. & Plyusnin, A. (2004). Oligomerization of Hantavirus N protein: C-terminal alpha-helices interact to form a shared hydrophobic space. *J Virol* **78**, 13669-77.
39. Alfadhli, A., Steel, E., Finlay, L., Bachinger, H. P. & Barklis, E. (2002). Hantavirus nucleocapsid protein coiled-coil domains. *J Biol Chem* **277**, 27103-8.
40. Alminaite, A., Halttunen, V., Kumar, V., Vaheri, A., Holm, L. & Plyusnin, A. (2006). Oligomerization of hantavirus nucleocapsid protein: analysis of the N-terminal coiled-coil domain. *J Virol* **80**, 9073-81.
41. Boudko, S. P., Kuhn, R. J. & Rossmann, M. G. (2007). The coiled-coil domain structure of the Sin Nombre virus nucleocapsid protein. *J Mol Biol* **366**, 1538-44.
42. Severson, W., Xu, X., Kuhn, M., Senutovitch, N., Thokala, M., Ferron, F., Longhi, S., Canard, B. & Jonsson, C. B. (2005). Essential amino acids of the hantaan virus N protein in its interaction with RNA. *J Virol* **79**, 10032-9.
43. Xu, X., Severson, W., Villegas, N., Schmaljohn, C. S. & Jonsson, C. B. (2002). The RNA binding domain of the hantaan virus N protein maps to a central, conserved region. *J Virol* **76**, 3301-8.
44. Wang, Y., Boudreaux, D. M., Estrada, D. F., Egan, C. W., St Jeor, S. C. & De Guzman, R. N. (2008). NMR structure of the N-terminal coiled coil domain of the Andes hantavirus nucleocapsid protein. *J Biol Chem* **283**, 28297-304.

45. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-93.
46. Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* **278**, 313-52.
47. Grzesiek, S. & Bax, A. (1993). The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements. *J. Am. Chem. Soc.* **115**, 12593-12594.
48. Grzesiek, S., Dobeli, H., Gentz, R., Garotta, G., Labhardt, A. M. & Bax, A. (1992). 1H, 13C, and 15N NMR backbone assignments and secondary structure of human interferon-gamma. *Biochemistry* **31**, 8180-90.
49. Wittekind, M. & Mueller, L. (1993). HNCACB, a high sensitivity 3D NMR experiment to correlate amide proton and nitrogen resonances with the α-carbon and β-carbon resonances in proteins. *J. Magn. Reson.* **101B**, 201-205.
50. Muhandiram, D. R. & Kay, L. E. (1994). Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. ser. B* **103**, 203-216.
51. Wishart, D. S. & Nip, A. M. (1998). Protein chemical shift analysis: a practical guide. *Biochem Cell Biol* **76**, 153-63.
52. Tolman, J. R., Chung, J. & Prestegard, J. H. (1992). Pure-phase heteronuclear multiple-quantum spectroscopy using field gradient selection *J. Magn. Reson.* **98**, 462-467.
53. Grzesiek, S. & Bax, A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J Biomol NMR* **3**, 185-204.
54. Fesik, S. W. & Zuiderweg, E. R. P. (1988). Heteronuclear three-dimensional NMR spectroscopy. A strategy for the simplification of homonuclear two-dimensional NMR spectra. *J. Magn. Reson.* **78**, 588-593.
55. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989). Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry* **28**, 6150-6.
56. Guntert, P. (2004). Automated NMR structure calculation with CYANA. *Methods Mol Biol* **278**, 353-78.
57. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham Iii, T. E., Wang, J., Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Tsui, V., Gohlke, H., Radmer, R. J., Duan, Y., Pitera, J., Massova, I., Seibel, G. L., Singh, U. C., Weiner, P. K. & Kollman, P. A. (2002). Amber7. *Amber7*, University of California, San Francisco.

58. Dames, S. A., Martinez-Yamout, M., De Guzman, R. N., Dyson, H. J. & Wright, P. E. (2002). Structural basis for Hif-1 alpha /CBP recognition in the cellular hypoxic response. *Proc Natl Acad Sci U S A* **99**, 5271-6.
59. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* **8**, 477-86.
60. Delano, W. L. "The Pymol Molecular Graphics System." Delano Scientific LLC, San Carlos, CA, USA. http://www.pymol.org/.
61. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-41.
62. Deyde, V. M., Rizvanov, A. A., Chase, J., Otteson, E. W. & St Jeor, S. C. (2005). Interactions and trafficking of Andes and Sin Nombre Hantavirus glycoproteins G1 and G2. *Virology* **331**, 307-15.
63. Choy, N., Raussens, V. & Narayanaswami, V. (2003). Inter-molecular coiled-coil formation in human apolipoprotein E C-terminal domain. *J Mol Biol* **334**, 527-39.
64. Kiss, R. S., Kay, C. M. & Ryan, R. O. (1999). Amphipathic alpha-helix bundle organization of lipid-free chicken apolipoprotein A-I. *Biochemistry* **38**, 4327-34.
65. Zhou, N. E., Zhu, B. Y., Kay, C. M. & Hodges, R. S. (1992). The two-stranded alpha-helical coiled-coil is an ideal model for studying protein stability and subunit interactions. *Biopolymers* **32**, 419-26.
66. Spiropoulou, C. F. (2001). Hantavirus maturation. *Curr Top Microbiol Immunol* **256**, 33-46.

SUPPLEMENTAL FIGURES

Table 8-S1. Detail of the NMR Experiments for His_6-N^{1-74}.

Name	Bruker pulsesequence	Details	Time*
2D HSQC	hsqcfpf3gpphwg	8 NS* 2048×128 (N) pts	20 min
2D HMQC	blhmqcnoesy2d	8 NS 2048×256 (C) pts	1h 30min
3D HNCA	hncagpwg3d	8 NS 2048×50(N)×100(C) pts	13h
3D CBCA(CO)NH	cbcaconhgpwg3d	8 NS 2048×50(N)×100(C) pts	14h
3D HNCO	hncogpwg3d	16 NS 2048×50(N)×50(C) pts	7h
3D HNCACB	hncacbgpwg3d	32 NS 2048×50(N)×100 (C) pts	2d 6h
3D HBHA(CO)NH	hbhaconhgpwg3d	12 NS 2048×50(N)×96(H) pts	19.5h
3D HSQC-NOESY	noesyhsqcf3gp193d	8 NS 2048×50(N)×256(H) pts	1d 14h
3D HMQC-NOESY	blhmqcnoesy3d	8 NS 2048×76(C)×248(H) pts	2d 14h

NS*: number of scans
Time*: total acquisition time

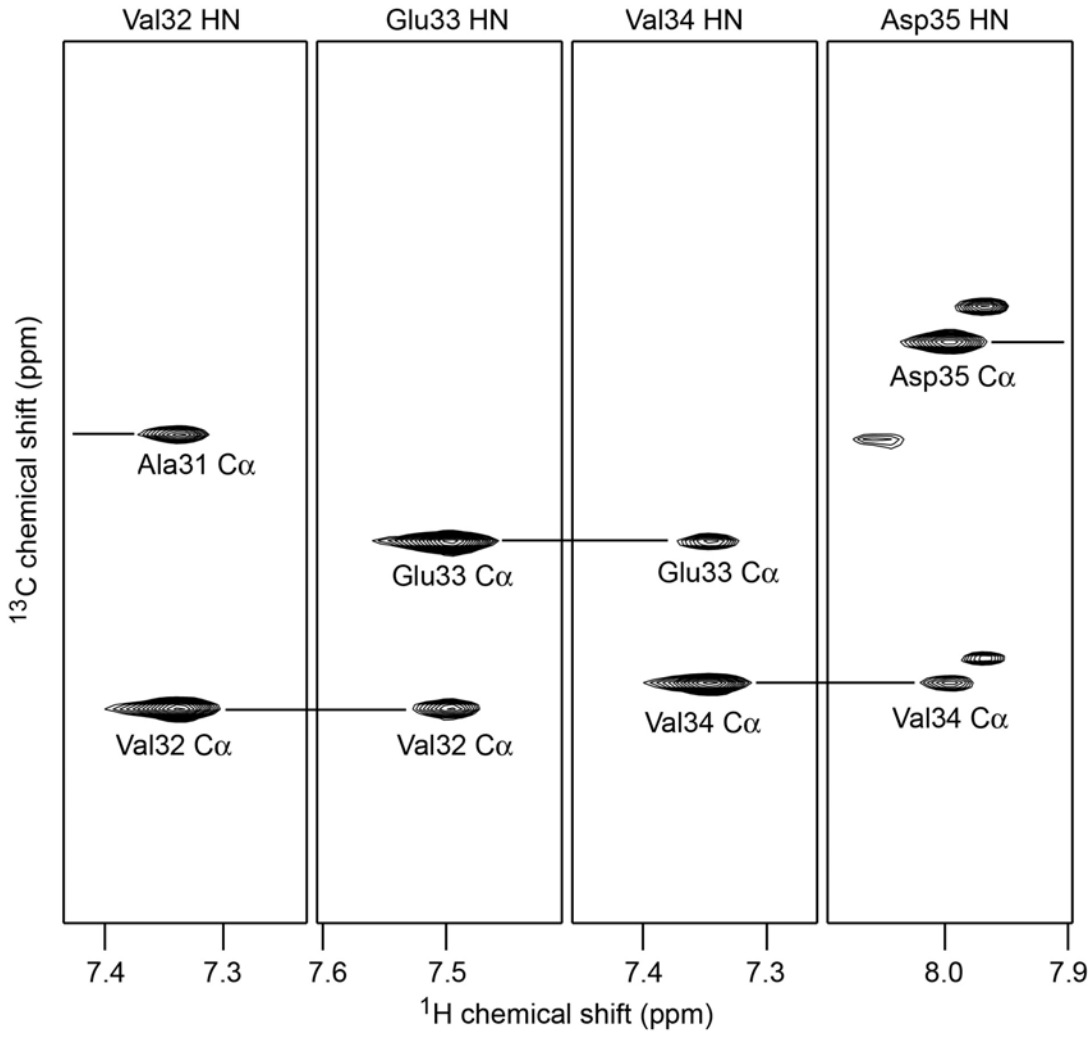

Figure 8-S1. Portion of the 3D HNCA strips of ^{13}C, ^{15}N-labeled 107-residue protein (His_6-N^{1-74}), shown as ^{13}C, ^{1}H strips for Val32-Asp35 (the ^{15}N ppm is on the Z-axis). Each strip shows the C_α of the preceeding and current residue and the amide (^{1}H and ^{15}N) resonances of the current residue, indicating sequence specific connectivity between these strips. The HNCA data is used together with ^{15}N-HSQC-NOESY, CBCA(CO)NH and HNCACB to accomplish the complete backbone assignment.

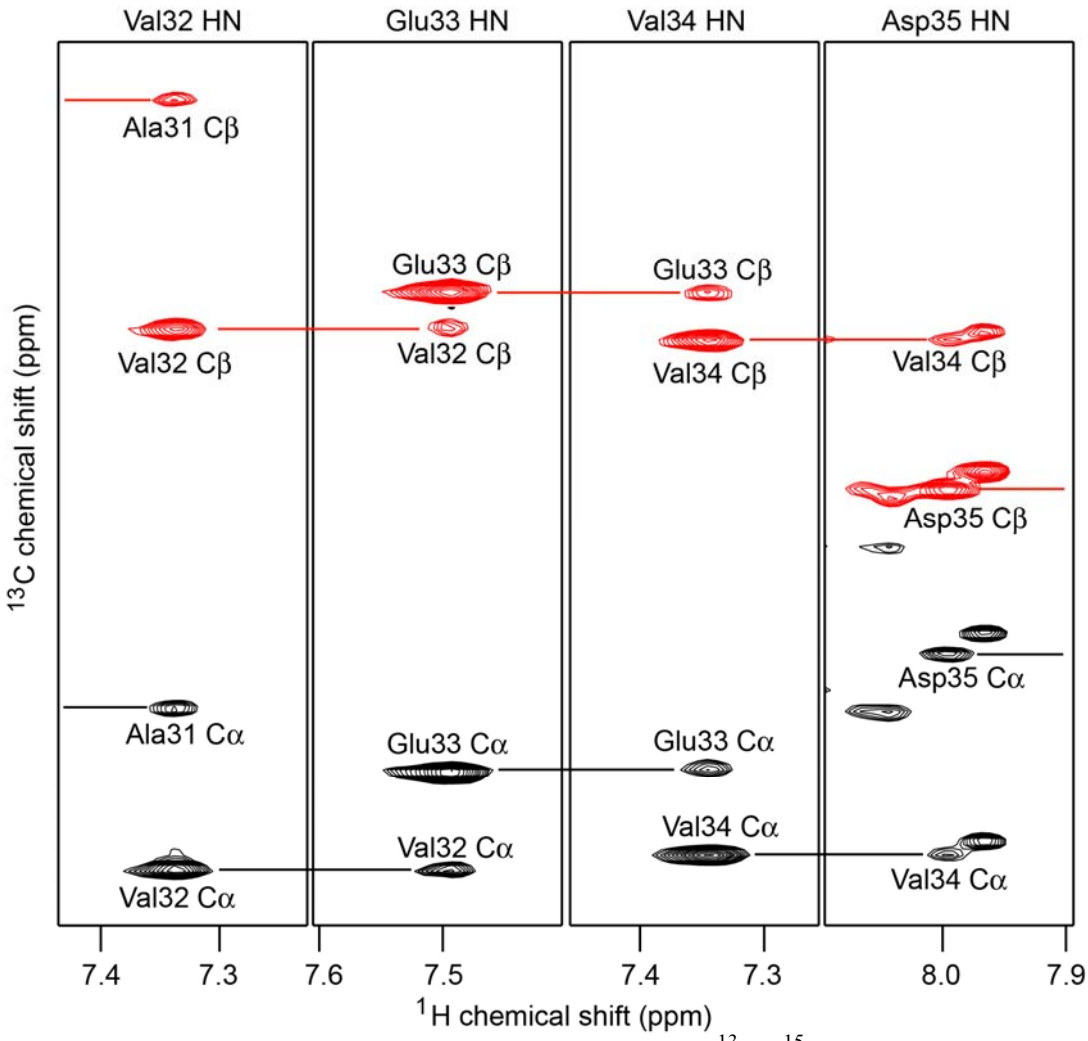

Figure 8-S2. Portion of the 3D HNCACB strips of ^{13}C, ^{15}N-labeled 107-residue protein (His_6-N^{1-74}), shown as ^{13}C, ^{1}H strips for Val32-Asp35 (the ^{15}N ppm is on the Z-axis). Each strip shows the C_α and C_β of the preceeding and current residue and the amide (^{1}H and ^{15}N) resonances of the current residue, indicating sequence specific connectivity between these strips.

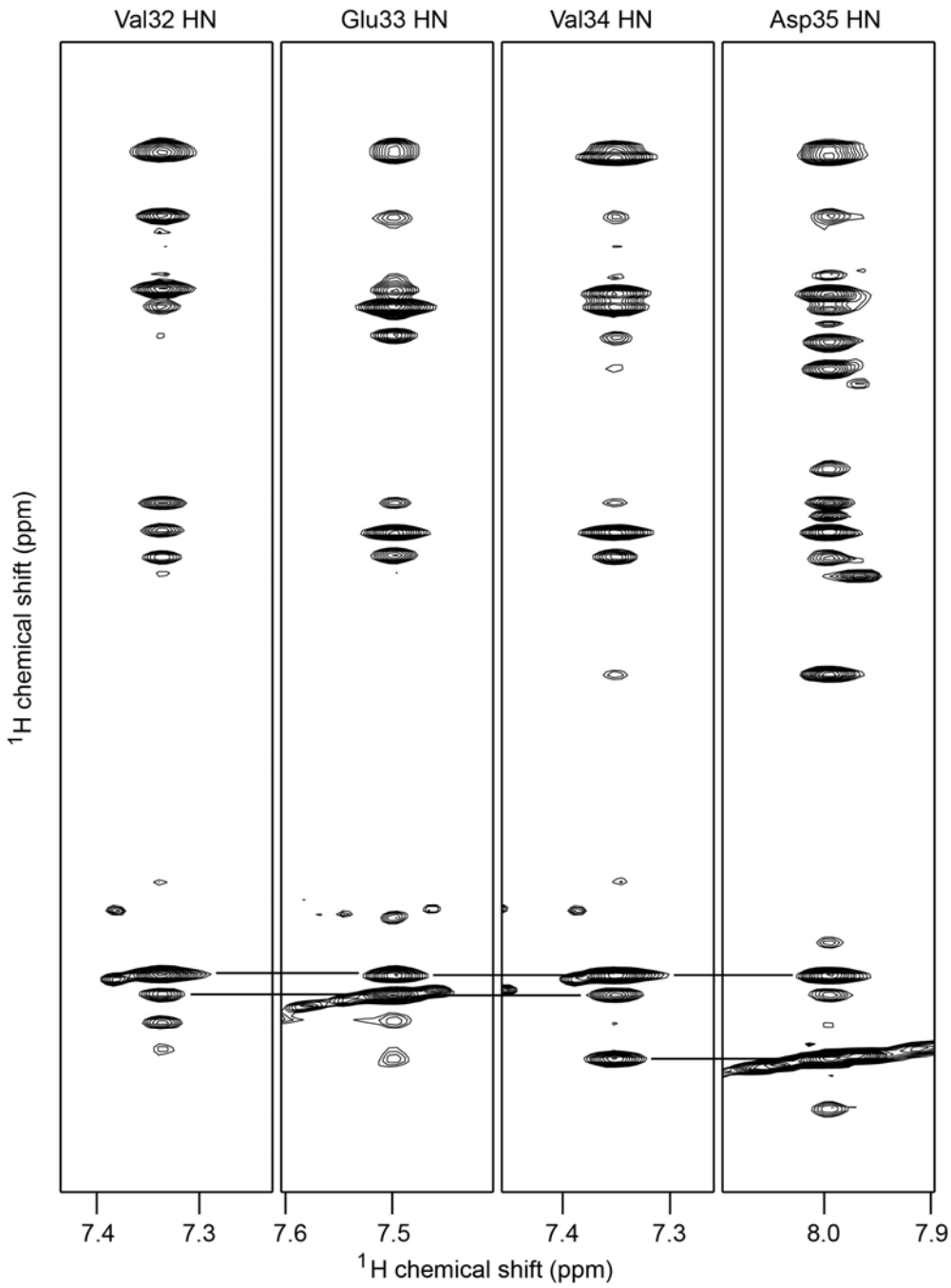

Figure 8-S3. Portion of the 3D ^{15}N-HSQC-NOESY strips of ^{15}N-labeled 107-residue protein (His_6-N^{1-74}), shown as ^{1}H, HN strips for Val32-Asp35 (the ^{15}N ppm is on the Z-axis). Each strip shows all the backbone H_N NOE peaks including its own H_N and the H_Ns of the preceding and following residues, indicating a sequence specific connectivity between these strips which can facilitate the backbone assignments.

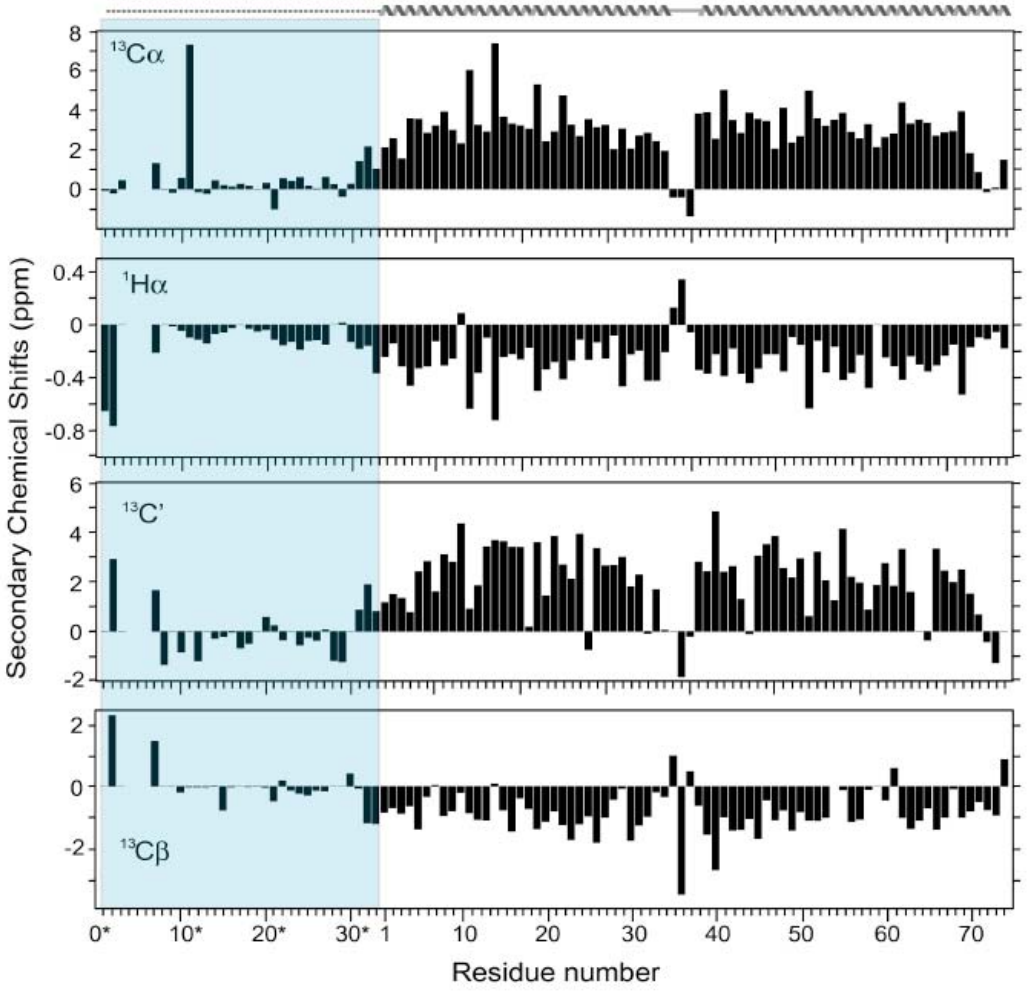

Figure 8-S4. Secondary C_α, H_α, C', and C_β chemical shifts show that the first 33 residues (shaded), which are the His-tag from pET151, lack secondary structure whereas the Andes virus $N^{1\text{-}74}$ region consists of two α-helices.[44]

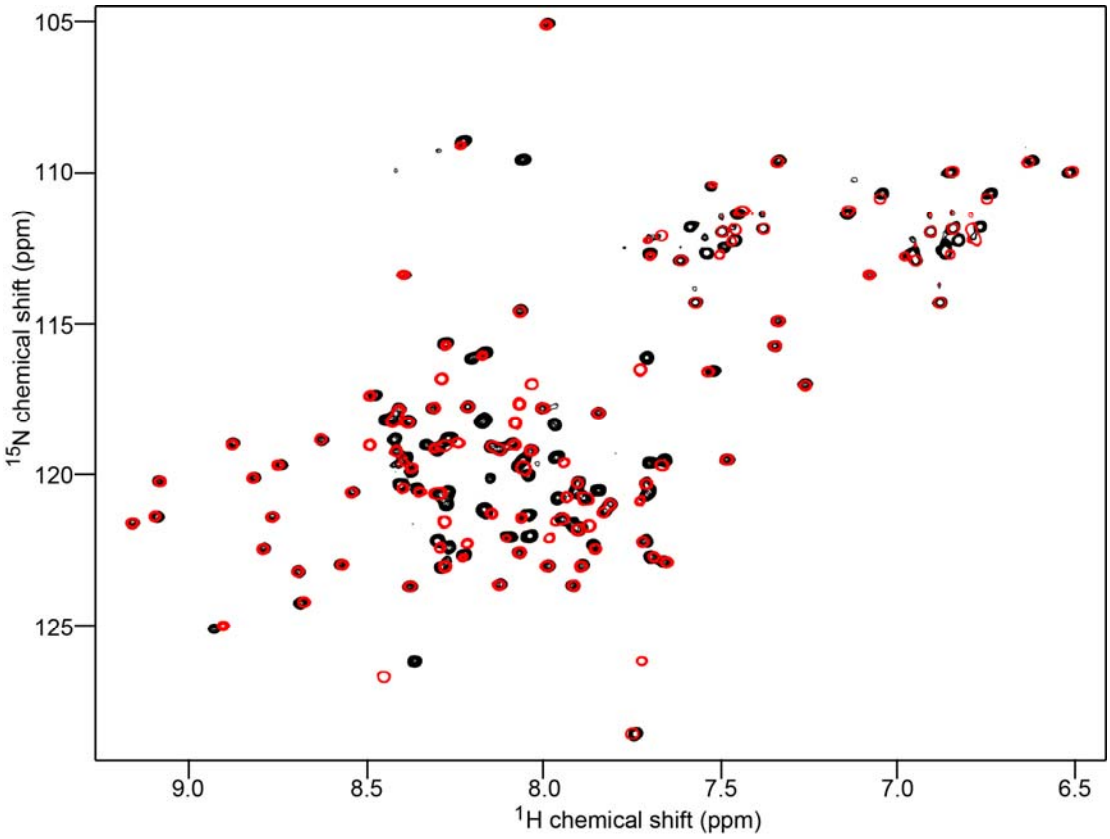

Figure 8-S5. Overlay of two 2D HSQC spectra of ^{15}N-His$_6$-N$^{1-74}$ and the TEV-digested ^{15}N-N$^{1-74}$. Most of the peaks are overlapped, indicating the extra His$_6$ did not affect the coiled-coil structure.

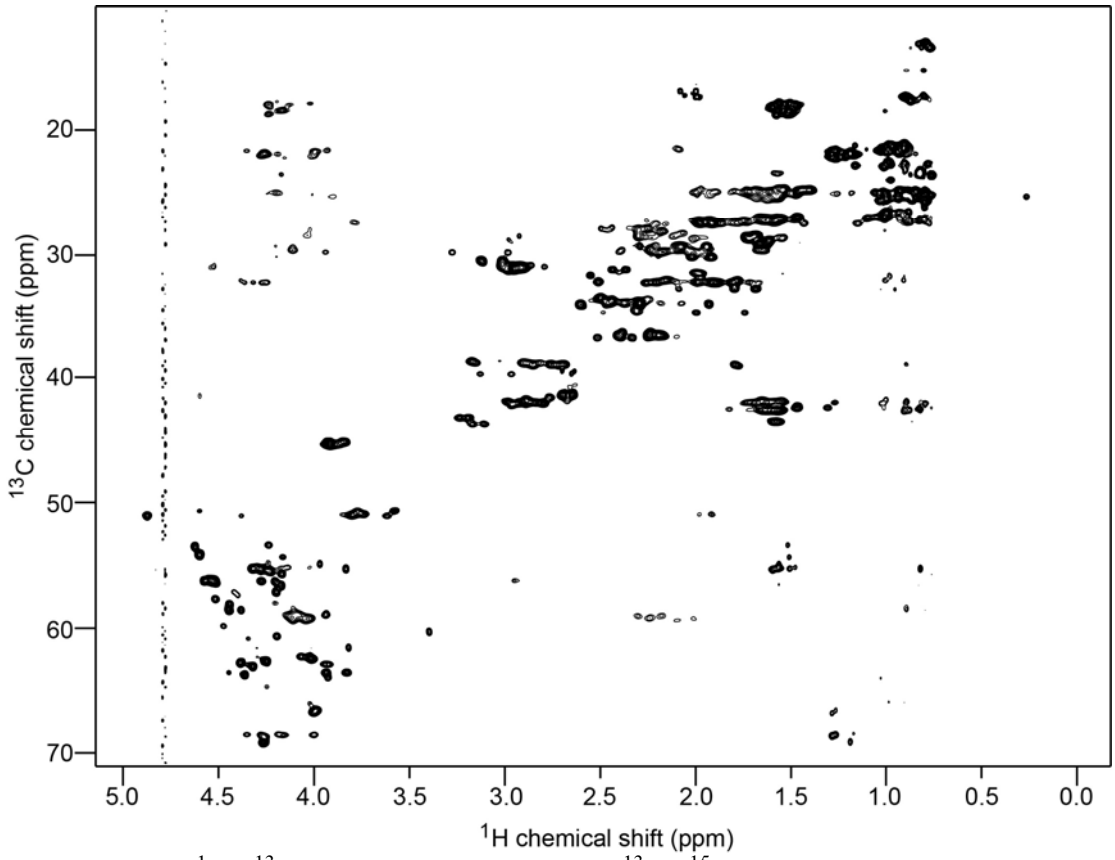

Figure 8-S6. ^{1}H, ^{13}C-HMQC 2D spectra of ^{13}C, ^{15}N-labeled 107-residue protein (His_6-N^{1-74}) in 100% D_2O. The 3D ^{13}C-HMQC-NOESY experiment was based on this 2D HMQC experiment and a third ^{1}H NOE dimension was added. With the help of backbone assignment and 3D HBHA(CO)NH, complete assignment of 2D ^{1}H, ^{13}C-HMQC and 3D ^{13}C-HMQC-NOESY of His_6-N^{1-74} were achieved.

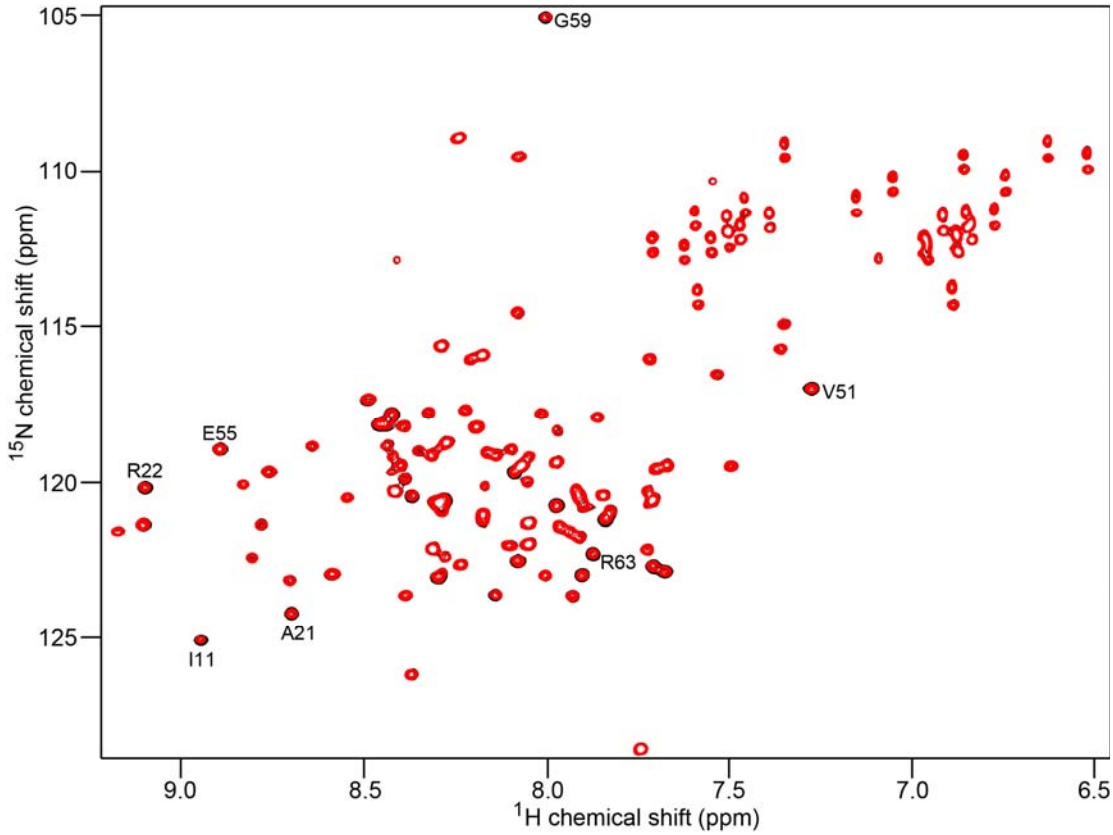

Figure 8-S7. Overlay of two His_6-N^{1-74} HSQC spectra showing the hydrogen-deuterium exchange experiment. Residues I11, A21, R22, V51, E55, G59 and R63 with slower hydrogen-deuterium exchange rate were pointed out to show the typical volume differences due to the protection of hydrogen bonds in α-helices. HSQC of black peaks was acquired at 5 mins after dissolved in 50% H_2O/50% D_2O; HSQC of red peaks was acquired at 16 hours later.

Figure 8-S8. Sequence alignment of hantavirus nucleocapsid N^{1-74} coiled-coil domain with the conserved hydrophobic (gray) and polar (yellow) heptads highlighted. The sequences are arranged according to hantaviral species that cause: **(top)** HCPS (Hantavirus CardioPulmonary Syndrome), **(middle)** nonpathogenic or mild form of HFRS (Hemorrhagic Fever with Renal Syndrome) and **(bottom)** severe form of HFRS. The hantavirus species are: AND, Andes; SNV, Sin Nombre; LAN, Laguna Negra; MUL, Muleshoe; BAY, Bayou; RIO, Rio Mamore; ELM, El Moro Canyon; NYV, New York; TUL, Tula; ISL, Isla Vista; PHV, Prospect Hill; PUU, Puumala; TOP, Topografov; KHA, Khabarovsk; SEO, Seoul; THA, Thailand; HTN, Hantaan; DOB, Dobrava.[44]

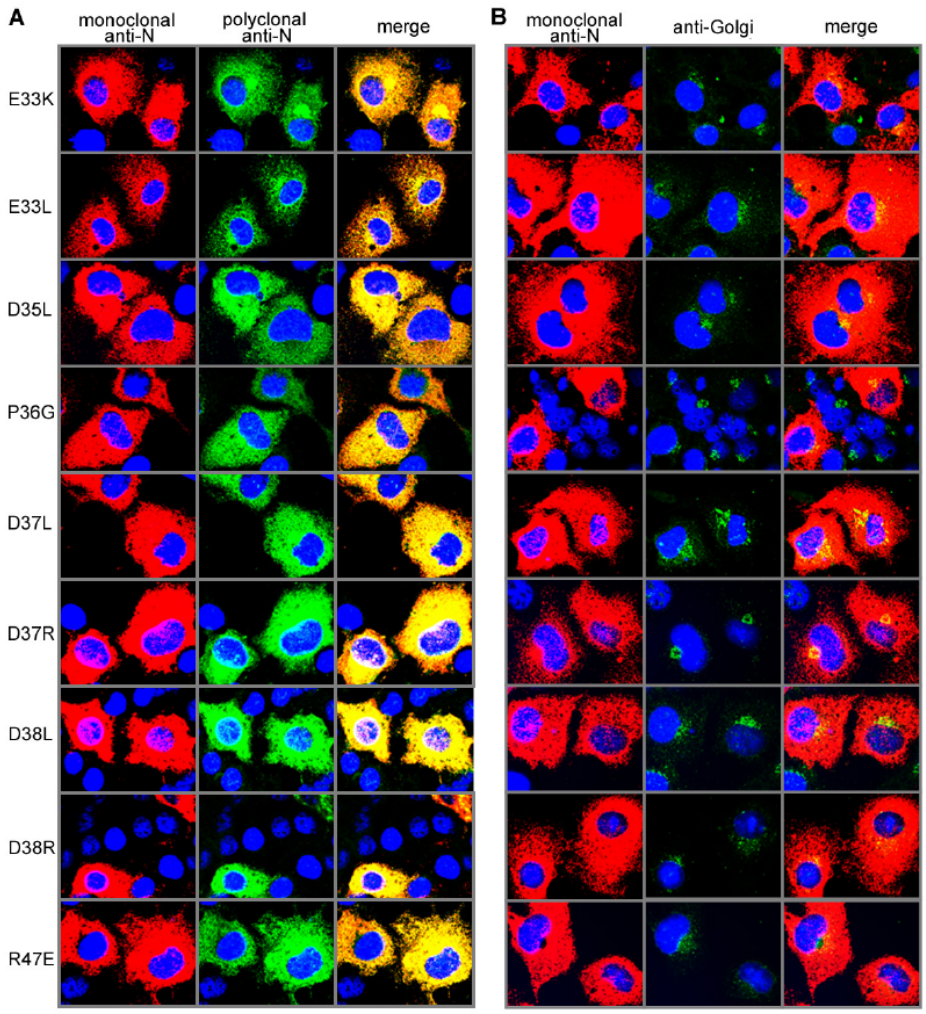

Figure 8-S9. Immunocytochemistry of full length N protein with mutations in the N^{1-74} coiled-coil domain. Cos-7 cells were transfected with will type and mutant N protein, and doubly stained with **(A)** monoclonal and polyclonal anti-N antibodies, and **(B)** monoclonal anti-N and Golgi-specific antibodies. The anti-N monoclonal antibody was AB34757 from Abcam (Cambridge, Mass.).[44] (Data courtesy of Daniel M. Boudreaux from Dr. Stephen C. St. Joer, University of Nevada, Reno)